



Is Act-Utilitarianism Self-Defeating?

Peter Singer

The Philosophical Review, Vol. 81, No. 1 (Jan., 1972), 94-104.

Stable URL:

<http://links.jstor.org/sici?sici=0031-8108%28197201%2981%3A1%3C94%3AIAS%3E2.0.CO%3B2-8>

The Philosophical Review is currently published by Cornell University.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/sageschool.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

IS ACT-UTILITARIANISM SELF-DEFEATING?¹

THE normative principle that all acts are to be judged by their consequences—the principle of act-utilitarianism—has been subjected to a great deal of criticism, but continues to have adherents, of whom, I may as well say straightaway, I am one. Most of the criticism has been inconclusive because it has consisted of the outlining of unusual situations, in which the application of act-utilitarianism is said to give results which conflict with our “ordinary moral convictions.” This method of argument can never move anyone who has greater confidence in the act-utilitarian principle than in his “ordinary moral convictions.” Whenever the conflict is a real one, and not merely an apparent conflict, dependent on the omission of factors which the act-utilitarian can and should take into account, the genuine act-utilitarian will be prepared to jettison his “ordinary moral convictions” rather than the principle of act-utilitarianism.

The argument of Hodgson's *Consequences of Utilitarianism* is challenging precisely because it avoids this common approach. Hodgson recognizes the inconclusive nature of previous anti-utilitarian arguments. He advances a different kind of argument which, he confidently asserts, is capable of showing convincingly that the principle of act-utilitarianism is not a rational ethical principle. This makes the central argument of his book worthy of detailed consideration. This central argument is intended to show that

to act upon the act-utilitarian principle would probably have worse consequences than would to act upon more specific moral rules, quite independently of the misapplication of that principle [p. 3].

In fact, if this were true, it would still not quite refute act-utilitarianism. The two statements (i) “An act is right if and only if it would have best consequences” and (ii) “If people accepted (i) it would not have best consequences, even if they applied it correctly” are not inconsistent. Nevertheless, to establish the truth of (ii) would be seriously to embarrass the act-utilitarian, for he could hardly continue to advocate his doctrine. With this in mind, let us consider Hodgson's argument.

¹ D. H. Hodgson, *Consequences of Utilitarianism*. (Oxford, Clarendon Press, 1967.) Page references in the text are all to this book. In developing the views expressed in this review I have been assisted by discussions with various people, especially Professor R. M. Hare, Mr. D. Parfit, and Professor H. R. West.

Hodgson makes his attack on act-utilitarianism in the second chapter of his book. This is the crucial chapter. The remainder of the book is largely a working out of the application of the argument of this chapter with regard to, first, rule-utilitarianism, and second, the use of utilitarianism in the justification of legal decisions. For this reason I shall, in this discussion, concentrate on the argument of Chapter II.

Hodgson's formulation of the principle of act-utilitarianism, which I shall accept for the purposes of discussion, is as follows:

An act is right if and only if it would have best consequences, that is, consequences at least as good as those of any alternative act open to the agent [p. 1].

Hodgson, for convenience, uses "best" to mean "best or equal best," and I shall do the same.

In the second chapter, Hodgson considers the consequences of acting on the principle of act-utilitarianism in two different situations: when all the members of a society accept the principle, and when one is an individual acting in a society consisting mainly of people who are not utilitarians. The crucial points of his argument, however, are all raised by the case of a society in which act-utilitarianism is universally accepted, and so, for the sake of brevity, I will restrict what I have to say to this case.

Before presenting his arguments, Hodgson specifies carefully the circumstances of the society we are to consider. These circumstances represent, according to Hodgson, an "act-utilitarian's ideal." They are important, not just for Hodgson's own arguments, but also for possible counterarguments. I shall quote them in full.

[L]et us consider a society in which everyone accepts the act-utilitarian principle as his only personal rule, and attempts always to act in accordance with it. We assume that everyone is highly rational, sufficiently so to understand the implications of the use of act-utilitarianism (including those to be demonstrated in this section). We assume too that the universal use of act-utilitarianism and universal rationality is common knowledge, in the sense that everyone knows of it, and everyone knows that everyone knows, and so on. We leave open the possibility that everyone might always succeed in acting in accordance with his personal rule. We assume that there are no conventional moral rules in this society: everyone knows that everyone else attempts with high rationality to act in accordance with act-utilitarianism, and so no-one is concerned to criticise the conduct of others or to make demands of them [pp. 38-39].

Hodgson's arguments concern keeping promises and telling the truth. To take promise-keeping first: when we ask why we should be

more concerned to do something we have promised we would do, than to do an act which we just happen to have mentioned we might do, the standard act-utilitarian reply is that the person to whom we made the promise normally has expectations of the promised act being performed, which he would not have if the act had merely been mentioned as a possibility. It is, ultimately, because of these expectations that the performance of the promised act will have greater utility than the performance of the act which was mentioned as a possibility. But, Hodgson asks, would this be true in an act-utilitarian society of the kind specified? His answer is that it would not, because in such a society the promisee will know that the promise made to him will not be kept unless keeping it has best consequences. The fact that the act was promised will not lead to its performance having greater utility than it would have had, had it not been promised, unless the promisee will, because of the promise, have a greater expectation of its being performed than he would otherwise have had. The promisee will have good reason for this greater expectation only if he believes that the promisor believes that the act will be expected by him, the promisee, with greater expectation than it would have been, had it not been promised, but merely mentioned. The promisor will know this, and the promisee will know that he knows, and so on. A spiral has been set up which cannot be cut across. Any attempt to build up a basis for a greater expectation of the promised act is, Hodgson says, mere bootstrap-tugging. The expectation can have no rational basis, and hence there is no greater utility in doing something one has promised to do than there is in doing something one has merely mentioned one might do. So promising would be pointless in an act-utilitarian society.

A parallel argument applies to telling the truth. Imagine that *A* tells *B*: "*X* is *Y*." In an act-utilitarian society, *B* would have good reason to believe *A* only under the following conditions. If *X* were *Y*, it would, in *A*'s belief, be best to tell *B*; if *X* were not *Y* it would, in *A*'s belief, not be best to tell *B* that *X* is *Y*. These conditions will hold generally only if *B* is likely to take the information conveyed as true, for only then will the utilitarian benefits which come from the conveying of true information—such as the possibility of making arrangements based on the information—be possible. But as *B*'s taking the information to be true rather than false is a condition precedent of *A*'s having good reason to tell *B* the truth, the situation is precisely similar to that of promise-keeping.

Hodgson concludes that for these reasons a society in which everyone

acted according to act-utilitarianism would be at a grave disadvantage compared to a society in which people acted on moral rules. For without promise-keeping and the communication of information there would be no human relationships as we know them. Hodgson emphasizes that this conclusion applies even if everyone applies act-utilitarianism correctly in the circumstances in which they are, but these circumstances—universal acceptance of act-utilitarianism and highly rational application of it—are in fact highly unfavorable to the production of good consequences.

One question that might be asked about Hodgson's ingenious arguments is whether he has himself considered sufficiently carefully all the effects which the circumstances of the society he has described would have. It will be recalled that Hodgson specified that in this society everyone adopts the principle of act-utilitarianism as his only personal rule, and attempts always to act in accordance with it. Under these circumstances, people would not act from the motives which most commonly lead people to make false promises and to tell lies—motives like self-interest, malevolence, pride, and so on. Nor would there be any need to make false promises or tell lies from utilitarian motives, in the sort of circumstances of which critics of utilitarianism are so fond: there would be no need to make consoling promises to dying people who wish their estates to be distributed in some way contrary to utility, since dying people would not wish this; no need, either, to tell a lie to save a man from his would-be murderer. Hodgson fails to see that there is any problem here. He writes of act-utilitarians breaking promises or telling lies without suggesting how doing so would bring about best consequences. His argument is based not on the existence of a reason for lying or breaking a promise, but on the absence of a sufficient reason for telling the truth or keeping a promise. This is significant, as we can see if we try to construct an example.

Let us imagine a case in which *A* has the choice of telling *B* the truth or a lie. *A* and *B*, we shall say, are working together in an office. (In constructing an example, it is impossible to avoid begging the question at issue to some extent. If Hodgson is right in saying that in an act-utilitarian society no communication would be possible, then offices and the other elements of this example would not be possible either. If this is considered a weakness, we might avoid the difficulty by assuming that, an instant before the events of my example take place, everyone in an until-then-normal society is miraculously converted to act-utilitarianism.) On this particular day, *B* intends to work overtime. His only means of transport home is by bus. If he misses the

bus, he will have to walk, which will make him very tired, waste time, and lead to his wife's worrying about him. In this situation, of which both *A* and *B* are aware, *B* asks *A*: "What time does the last bus go?" *A* knows the answer. Is it not in accordance with act-utilitarianism for *A* to tell *B* the correct time? Hodgson would reply that it would have better consequences for *A* to tell *B* the truth only if *B* were likely to take the information as true, and *B* would know that *A* would have no reason to tell the truth unless *A* believed that he, *B*, was likely to take the information as true, and so on. But consider the matter from *A*'s point of view. He has the choice of telling *B* the correct time, a fictitious time, or saying nothing. There is no possibility, barring extraordinary accidents, of any beneficial consequences arising from any course of action except telling *B* the correct time; but there is a fifty-fifty chance that telling *B* the correct time will lead to the beneficial consequences of *B* going to the bus stop at the right time. For even if there is no good reason for *B* to believe that *A* will tell him the truth, there is also no good reason for him to believe that *A* will tell him a lie, and so there is an even chance that *B* will take the information *A* gives him to be true. It is of course possible that if *A* tells *B* a fictitious time, *B* will treat this false information as false, but this cannot ensure, or even make it likely, that *B* will go to the bus stop at the right time. The point here is just that there is only one way for *A*'s statement to be true, but many ways for it to be false. Because of this, *A* has a reason for telling *B* the truth.

Once there is some reason for *A* to tell the truth, there is more than enough reason for him to do so. For *B*, being highly rational, will have thought of the considerations just pointed to, and will be aware that there is a reason for *A* to tell him the truth, and *A* will know this, and so on. So we get the Hodgson spiral working in the other direction, and *A* will have the normal utilitarian reason for telling the truth—that is, that *B* will take the information to be true and make arrangements based on its truth.

It might be objected that I have constructed an especially favorable case. In a real-life situation, would it not be possible that a lie would have best consequences? In the example, for instance, might it not be the case that *A* believes that great good will come if *B* works an hour longer than he would if he left to catch the last bus? If this is possible, would not *A* be right, on act-utilitarian grounds, to tell *B* that the last bus left an hour later than it really does leave?

This objection forgets that both *A* and *B* take act-utilitarianism as their personal rule and always try to act on it. So if it is the case that

the good of *B* working an extra hour outweighs the disutility of his having to walk home, all that is necessary to ensure that he does the extra work is that *A* explain this to him, thereby avoiding at least some of the disutility that would come from *A* telling *B* that the bus comes later than it really does—*B* will not have to wait unnecessarily, and he can telephone his wife so that she will not worry. So *A* still has no good reason for lying.

A different objection to my example might be that it depends on a question being asked to which there is only one true answer, and more than one false answer. Does our conclusion apply to other situations as well? In reply to this, one could say that it would seem to be possible to ask even ordinary questions, which would normally require a simple yes/no answer, in such a way as to make two false answers possible. If an office worker wished to know whether or not to reply to a letter, he could ask: "Shall I reply to this letter, file it, or make a paper dart out of it?" In this way the person addressed has a better chance of producing best consequences by saying what he really thinks best. Admittedly, if this were really necessary, act-utilitarianism would cause inconvenience, but it would not be disastrous, and it is certainly not clear that this inconvenience would outweigh the benefits of everyone's adopting act-utilitarianism.

In any case, there are other grounds for believing that in a society of act-utilitarians there would be sufficient reason for telling the truth in normal situations. Let us consider an example in which information is volunteered. I am walking along the street when *A* comes up to me and says: "There is a very good film on at the local cinema this week." How am I to take this remark? Is it possible that *A* wants me to go to the cinema for some reason, even though the film is very bad? Perhaps the cinema will have to close if it does not get good audiences this week, and the disutility of this outweighs the disutility of people being bored by the film. But this explanation will not do, for, as in the previous example, *A* could explain these facts to me, and I could buy a ticket without wasting my time by actually sitting through the film. Nevertheless, Hodgson might say, I cannot assume that *A* was telling me the truth. He may have been trying to warn me away from a very bad film, believing that I would take what he had told me to be false. This is not feasible either. Why would *A* have bothered to speak at all, since I am just as likely to take his remark to be true as to be false? Hodgson may claim that this is just his point. No one would have any reason to speak, and communication would cease. Before we accept this, however, consider the situation from the point of view of the

recipient of the information. Since by going through the business of inverting what *A* says to me—thinking to myself, “He says the film is good, but he may be telling a lie, so the film may be bad”—I am no more likely to arrive at the truth than if I take what *A* says at face value, why should I bother to invert it? Am I not just a fraction more likely to take it at face value? If I am, *A*, being highly rational, will know this, and will know that he is more likely to produce best consequences if he tells the truth, while I, being highly rational, will know this, and so expect *A* to tell the truth . . . and so we get the spiral unspiraling once again, and we have all the reason we need for telling the truth.

Analogously with the argument just made, we could also ask why *A* should not save himself the bother of inventing a lie by telling the truth, thus making it fractionally more likely that he would tell the truth, and reversing the spiral once again. Hodgson attempts to forestall this objection by saying that any disvalue involved in the need to invent a lie would be balanced by the satisfaction of exercising the skill of lying (p.43). Hodgson apparently has not noticed that the point is equally effective if made in regard to the recipient of the information, and his reply, which is not particularly convincing in the case of the person making the statement, would be quite implausible if made in respect of the recipient.

Hodgson does at one point suggest that even if it were possible to arouse expectations in the recipient that the information is true, it would not be possible to place much reliance on it, because it would still be better to tell a lie if the consequences on the whole would be better—and since the recipient would know this, he would not have very strong expectations (p. 50). This again seems to overlook the fact that if everyone were an act-utilitarian most of the reasons, selfish and unselfish, which we would otherwise have for lying would not exist. Hence I believe that once the expectations can be aroused, at least as much reliance could be placed on them as is possible in our society at present, outside the circle of those we know to be sincere.

I have questioned only Hodgson’s argument about truth-telling, but similar points could be made about his argument in respect of keeping promises. If there is little reason for making false statements, then there is little reason for making false statements of intention. But a promise implies, in some sense, a statement of intention, and whatever the promise adds to the statement of intention would not seem to affect the validity of the application of the previous argument about statements in general to the statements of intention implied by promises. In fact, it seems to me that a statement of firm intention to do an

act, coupled with a recommendation to the hearer to make arrangements based on the expectation that the intention be carried out, is just as useful as, if it is not equivalent to, a promise. If, because of unforeseen events, the "promisor" is in doubt as to whether doing as he said he intended to do will have best consequences, he must, as an act-utilitarian, take into account the expectations raised and arrangements which may have been made as a result of his statement of intention. This, of course, is as much as an act-utilitarian would ever want to say in defense of the institution of promising.

Quite apart from these objections to Hodgson's central argument, there is a more obvious one, which he does consider but not, in my opinion, refute. It is independent of the arguments I have put so far, and for the purposes of discussing it, we may assume that what I have said up to now has been mistaken.

The obvious objection is that if the situation were as Hodgson describes it, it would be justifiable on act-utilitarian grounds to take steps to form a social practice of telling the truth and making and keeping firm statements of intention (which I shall, for convenience, continue to call "promises"). Any steps toward the formation of these practices would have the good consequences of making desirable activities possible. Since telling the truth and keeping promises could help in the formation of these practices, while lying and breaking promises could not, this would give an additional reason for telling the truth and keeping promises. The spiraling effect would come into operation. This would ensure the rapid development of the practices. The informer or promisor would then have the dual reasons of preserving the useful practice and fulfilling expectations.

Hodgson seems to be aware of this kind of objection to his arguments. Yet his reply to it is puzzling:

Such steps could have good consequences, but, although perhaps justified by act-utilitarianism, they would amount to a partial rejection of act-utilitarianism and so would be inconsistent with our assumptions. These steps would amount to a partial rejection of act-utilitarianism, because the persons would be forming habits to do acts known not to be justified according to act-utilitarianism; and they could form these habits only if they resolved to refrain from applying act-utilitarianism in relation to these acts [p. 48].

I am puzzled by the statement that acts could be justified by act-utilitarianism, and yet amount to a partial rejection of act-utilitarianism. This looks like a contradiction. Perhaps Hodgson means that while the taking of steps to get the habit, or practice, established is

justified by act-utilitarianism, the practice itself is one of refraining from the calculation of consequences in respect of the particular acts, so that acts done in accordance with the practice may not be justified by act-utilitarianism. There are two points that may be made in reply to this. First, if acts may be justified because they help to get a practice established, surely they may also be justified because they help to preserve a useful, established practice. Second, Hodgson's admission that the acts which establish the practice may be justified by act-utilitarianism undermines the arguments he made earlier; for once the practice is established the point about lack of expectation, that promises will be kept and information given true, will not apply. Where there is a practice there are expectations, and the standard act-utilitarian justifications of keeping promises and telling the truth will operate.

It may be that in talking of "forming habits to do acts known not to be justified according to act-utilitarianism" Hodgson has in mind the formation of habits or practices of *always* telling the truth, and *always* keeping promises, no matter what the consequences. This would certainly be inconsistent with act-utilitarianism, but it would also be unnecessary. The benefits of communication and reliability may be gained without having such absolutist practices. All that is necessary is that there be habits of telling the truth and keeping promises unless there is a clear disutility in doing so which outweighs the benefits of preserving the useful practices and fulfilling the expectations aroused. It is, after all, an advantage of act-utilitarianism that it does not force us to reveal the hiding places of innocent men to their would-be murderers, or leave accident victims groaning by the roadside in order to avoid being late for an appointment we have promised to keep.

It might be more plausible to argue that it is the initial acts, before the practice has been established, and the expectations aroused, that would be contrary to act-utilitarianism. Hodgson does not argue this in the context of the passage we have been discussing, but in a subsequent discussion of the justification of a decision by a judge to punish an offender, Hodgson argues that although an unbroken record of punishment might deter potential offenders, such an unbroken record can never, on act-utilitarian grounds, get started. Hodgson's argument is that no single case can be a necessary or sufficient condition for such an unbroken record, because if we did not punish in any particular case, we could still have an unbroken record from the *next* case onward which would deter just as well [p. 93]. This argument seems to be

based on the assumption that the only consequences of an act which may be taken into account, in deciding whether that act is justified by the act-utilitarian principle, are those for which the act is a necessary or sufficient condition. (This assumption has, incidentally, been the basis of claims by other writers that act-utilitarianism cannot explain why we ought to vote at elections, or obey power restrictions, when failure to do so will not bring about the defeat of our candidate or a general power breakdown.) Although some act-utilitarian writers may have assumed that only consequences for which the act is a necessary or sufficient condition should be taken into account, there is no good reason for an act-utilitarian to do so. An act may contribute to a result without being either a necessary or sufficient condition of it, and if it does contribute, the act-utilitarian should take this contribution into account. The contribution that my vote makes toward the result I judge to be best in an election is a relevant consideration in deciding whether to vote, although it is, almost certainly, neither a necessary nor a sufficient condition of that result; for if this were not so, the act-utilitarian view would leave us with a result which was unconnected with the actions of any of the voters, since what is true of my vote is equally true of every individual vote. In the punishment case, the first act of punishing may be justified, on act-utilitarian grounds, by its probable contribution to an unbroken record of punishment which will have a deterrent effect. In the cases we were considering originally, an act of telling the truth or keeping a promise will normally have greater utility than would its opposite, because it has a reasonable chance of contributing to the beneficial consequences of setting up a desirable practice. Our act-utilitarians, being highly rational, would understand this, and so contribute to the establishing of the practice themselves, as well as expecting other act-utilitarians to do so. The expectations so generated would increase the utility of conforming to the practice, which would therefore become established very quickly.

It seems to me, then, that Hodgson fails to establish the challenging central thesis of his book; and as I have said, the remainder of the work is based on the arguments we have just been discussing. This does not mean that the later sections are without interest, once these arguments have been rejected. On the contrary, there is much here that is stimulating for anyone interested in rule-utilitarianism or the justification of legal decisions—particularly the latter topic, which takes up almost exactly half of the book's total length.

PETER SINGER

I should also say, perhaps, that in dissenting from Hodgson's conclusions I have not been concerned to deny that there are no problems at all, of the sort Hodgson raises, in being an act-utilitarian. There may be occasions when a person is handicapped by being known to be an act-utilitarian—for example, a doctor, who assures a seriously ill patient, depressed and fearful that he will die, that his condition is hopeful, is less likely to be believed if he is known to be an act-utilitarian than if he is known to believe that lying is always wrong. These occasions would, I think, be few enough and unimportant enough for the balance of advantage to favor act-utilitarianism. My concern has been to show that act-utilitarianism does not have the catastrophic consequences which Hodgson argues it would have.

PETER SINGER

University College, Oxford