

# Moral Argument

Jonathan Bennett

from: *Mind* 69 (1960), pp. 544–9.

[Added in 2012: The central thesis of this rather modest piece of work is illustrated with overwhelming brilliance and accuracy by Mark Twain in a passage that is reported and discussed in [www.earlymoderntexts.com/jfb/huckfinn.pdf](http://www.earlymoderntexts.com/jfb/huckfinn.pdf)]

The thesis is advanced by R. M. Hare that a judgment on an action or state of affairs is a *moral* judgment only if the person who makes it accepts some universal moral principle which, together with some true statement about the non-moral characteristics of the situation originally judged, entails the original judgment.<sup>1</sup> Instances of this thesis would take some such form as saying that someone who says ‘You ought not to have done what you did’ cannot be expressing a moral judgment by this unless he accepts something of the form ‘Actions which are A are wrong and what you did was A’ or ‘People who are A ought not to perform actions which are B, and you are A and what you did was B’. Hare claims that it is analytic that every moral judgment is so supported; he claims, that is, that ‘universalisable’ is part of the meaning of ‘moral’.

I think that Hare is perfectly right about this, but the question of truth is not what I am now primarily concerned to discuss. What I wish to bring out here is something of what can be built on the basis of Hare’s thesis, for I think that it is important in ways which Hare has not publicly

discussed. Furthermore, there is something to be said for the view that the best way to argue for the truth of the thesis is to bring out clearly what makes it important.

It is sometimes urged that the thesis is true but trivial. It is said: ‘If Hare’s thesis is that my judgment qualifies as a moral one only if it is an application of some universal moral principle which I accept, then the thesis allows any *prima facie* case of a moral judgment to qualify as a genuine moral judgment: nothing could fail to qualify. For it is always possible to form a universal principle of the form “Anything which is [*complete* description of the subject of the original judgment] is [moral characterisation as given in the original judgment]”, and to claim honestly enough that one accepts this “universal principle”—to claim, indeed, that accepting the original judgment *was* accepting the ‘universal principle’. This is just a logical trick, admittedly, but does the thesis admit of it, and if not how does it exclude it?’

Hare, to whom I am indebted for a discussion of these matters, argues that even when the thesis of universalisability is construed in such a weak form as not to exclude

---

<sup>1</sup> R. M. Hare, ‘Universalisability’, *Proceedings of the Aristotelian Society* 55 (1954–55).

this 'logical trick', it has consequences which are far from trivial. His tracing out of these is of the greatest interest, and one must hope that it will soon appear in print. But my present concern is with what follows from the thesis if it is strengthened a little—just enough to meet the objection cited above. Such a strengthening could be achieved by an expansion of the first sentence of this paper to read '. . . together with some true statement about *some but not all* of the non-moral characteristics of the situation originally judged. . .'. In what follows, the universalisability thesis will be understood in this way.

The thesis as thus reformulated still meets with an objection (which runs to the end of this paragraph): The thesis now specifies that for a judgment to be a moral one it must be an application of a principle which is not just universal in the sense that it contains an initial universal quantifier followed by nothing but quantifiers, variables, predicates and logical connectives; but which is also genuinely general, in the sense that its acceptance would commit one to the same judgment on other individual situations which were not the same in all respects as the one dealt with in the original judgment. But in how many respects might they be alike? Are we to suppose that the thesis leaves it open, as in the present formulation it certainly *does* leave it open, for the universal principle to be so comparatively ungeneral that no such principle covers two actual situations? If the thesis is not to be as liberal as this, how can it specify its illiberality? On the other hand, if the thesis is to be regarded as verified by universal principles of such a degree of ungenerality, then it is 'trivial' not in the technical sense but in the equally damaging sense of being without interest or importance. For it claims to state a necessary condition of a judgment's being a moral one; but how can it possibly matter whether or not a judgment fulfils *this* condition? So far as this condition

is concerned, a judgment can be a moral one without its involving any commitment at all to any moral judgment on any other actual situation—past, present or future. If I make a moral judgment in saying 'What you did was wrong', then all that the universalisability thesis says is that I am committed to something of the form 'Any action which is. . . is wrong' where the blank can be filled with a description of your action which, while it stops short of completeness, is nevertheless so detailed that there is no chance at all that I shall ever again be called upon to act on it or judge in accordance with it.

The best way to meet this challenge is not to attempt a further tightening up of the universalisability thesis, but rather to show that the thesis has important consequences even when it is construed in this fairly weak form.

Suppose, then, that you and I accept the universalisability thesis, and that I pass a judgment of the form 'What he has just done was wrong', claiming that I intend this as a moral judgment. You do not agree with this judgment, and you argue with me about it, not in order to 'prove' that the judgment is 'false' but in order by rational means to get me to withdraw it, or perhaps, by rational means, to get those listening to the argument to dissent from my judgment. You challenge me to produce the universal moral principle which I accept and of which my judgment is an application. That is, you ask 'Why was it wrong?', insisting upon a certain sort of answer which I agree that I am committed to being able to give. If my reply to this is average, it will be fairly brief: 'Because what he did was A, and all A actions are wrong.' You are now in a position to proceed against me through my universal principle, not just through my specific judgment. 'All A actions are wrong? *All*? But suppose. . . ' and you now try to find a counter-instance to my principle; that is, to describe an action, real or imagined, which has

the property A but which I do not judge to be wrong at all. If A is a fairly uncomplex property, you will have little difficulty in finding such a counter-instance; for most people do not accept in their full generality any universal principles of the form 'Everything which is A is M', where 'A' stands for a brief non-moral description and 'M' stands for a moral characterisation. We may suppose, then, that you do succeed in finding a counter-instance to my principle. I accordingly admit that I mis-formulated my principle, and that I should have said 'All actions which are A and B are wrong', where the action originally judged was B and the action described in the counter-instance was not B.

You will be acting unfairly if you accuse me of dishonesty. The gambit which begins 'What I really meant to say was. . .' is often rightly suspect, but not here. It is possible genuinely to accept a moral principle and yet to have to struggle to formulate it; and accordingly, in very many situations in which I announce my principle in terms of A and then in the face of a counter-instance go back and expand it in terms of B, it would be proper to grant that I am in a straightforward sense getting nearer to the principle which I did in fact accept when I made the original judgment. In some cases, certainly, it will not be plausible to say that I really accepted the revised principle all along. Eyebrows might rightly be raised at a remark of the form 'I see now that I have always accepted the principle that any action which is A, B, C, . . . X, Y and Z is wrong; and that is why I condemned his action'. In such a case, the principle is so long and complex that it may be utterly implausible to say that it was in any plain sense (unformulated but) 'accepted' and that the specific moral judgment was 'derived' from it. But the universalisability thesis would be wrong if it demanded that the universal principle which backs any moral judgment must be one from which the maker of the

judgment can honestly claim to have derived his judgment. It suffices—and Hare clearly intends the thesis in this way—that anyone who makes a specific moral judgment shall under challenge be able to produce a universal principle which has the appropriate logical relations with the specific judgment and about which he can honestly claim **(a)** that he does now accept the principle, and **(b)** that he does now regard the specific judgment as one which he makes only because he accepts the principle. We might sum this up by saying that the universalisability thesis does not demand that a system of moral judgments be reached by rational steps, but it does demand that any specific moral judgment be, at least in theory, rationalisable on demand.

We may take it, then, that you will not deny my right to meet the challenge presented by your counter-instance, by expanding or weakening or complicating my universal principle. If you wish to continue the argument, you can now try to find a counter-instance to the expanded principle. Again, if 'A' and 'B' are taken to stand for reasonably average descriptive adjectives, the finding of a new counter-instance should not be too difficult. In the face of such a new counter-instance, I can go back to expand my universal principle still further, into something of the form 'Any action which is A, B, and C is wrong', where C is a property possessed by the action originally judged and lacked by the action described in the new counter-instance.

One important thing to note about this sort of argument is that I need never lose it. For in each step of the argument, I can adopt either one or two alternatives: I can deny that your would-be counter-instance is a counter-instance, saying 'Yes, that would be wrong too'; or I can admit that you have found a counter-instance, and revise my rule accordingly. In applying my principle to what you thought was a counter-instance to it, I may surprise you, but you cannot 'falsify' the moral

judgment which I thereby make. Nor can you 'falsify' any expanded universal principle which I may produce in the face of a genuine counter-instance.

But there is a second important point about this sort of argument, namely, that there is a sense in which you may indeed win, and I may lose, the argument. For I may very well find myself in a position where I am faced with a choice between **(a)** accepting a specific moral judgment on an action described in one of your would-be counter-instances, which it embarrasses me very much to accept, and **(b)** expanding my universal principle by adding to it some predicate which it embarrasses me very much to have to admit to be morally relevant. The point about the second alternative is that, while there is no need to be discomfited by the admission that one accepts a moral principle of the form 'Any action whose character is thus-and-so, whose motive is such-and-such, whose results are so-and-so, and whose agent has a black face, is wrong', it may well be embarrassing to admit that the reference to skin-colour needs to be included, that if it were omitted one would not accept the principle.

Some explanation is required of the term 'embarrassment' as I am using it here. My embarrassment may arise from my having to admit that I do really regard your counter-instance as a counter-instance, and to admit also that, as between the action described in the counter-instance and the action upon which I originally passed judgment, I can find no qualitative difference which I can honestly claim to regard as morally relevant. In such a case, if I am honest, I shall withdraw my original judgment. It will still be incorrect to say that you have shown my judgment to be false; but you will have shown that the price which has to be paid for its rationalisation is one which I am not prepared to pay; or, what amounts to the same thing, you will have shown that the totality of all my moral judgments cannot be rationalised at all, including

those like 'I must admit that I do not regard skin-colour as a morally relevant feature'. If the argument terminates in this way, I shall have genuinely learned something from it. I shall have learned that a specific moral judgment which I voiced perfectly sincerely is not tenable *by me*.

There is a second way in which I may be embarrassed by your argument. My difficulty may arise not because I am unable to square my conscience with either of the options with which you face me, but because the option which I am prepared to accept with no inner struggle (either the option of the form 'Well, I think that skin-colour does make a difference' or the option of the form 'You tell your story most touchingly, but I still think that your hero *did* act wrongly') is one which I know or think will be strongly dissented from by those who are listening, so that I am embarrassed by being forced into a position in which I can maintain my original judgment only at the price of losing support. It must not be thought that this sort of embarrassment is merely that of the man who has been forced to reveal something which he would rather conceal. Someone may, for example, believe that he can rationalise his condemnation of some action of a socialist government on the basis purely of moral principles which he believes all his audience to accept, but be led by argument to see and to admit that he cannot rationalise the judgment except by an appeal to peculiarly conservative principles which he knows not to be shared by all his audience. In such a case, his embarrassment will arise not from any revelation that he holds conservative principles, but rather from a demonstration that he must appeal to such principles in defence of the specific judgment which he has made on this occasion. Such a revelation or demonstration may hinge on his acceptance of a would-be counter-instance, or on his making a judgment of moral relevance. These two have their distinct logical functions

in moral argument, but no special sort of moral position is associated with either.

What all this comes down to is that if the universalisability thesis is correct then issues of consistency can be raised with respect to moral judgments on individual actions and situations. No two specific moral judgments with different subjects can be inconsistent as they stand. But a pair of such judgments, where one expresses condemnation and the other expresses permission or approbation, or where one expresses approbation and the other expresses 'permission-not-to' or condemnation, may be shown to be inconsistent in the moral system of the person who makes them. Specifically, they are inconsistent in this system unless the system contains universal moral principles from which both judgments follow, and also contains judgments of relevance such that the principle from which one judgment follows does not rule out the other judgment.

For an expression of condemnation to be inconsistent with an expression of permission-or-approbation is simply for it to imply another condemnation; and this fact provides a way of re-stating the matter of the previous paragraph so as to bring it closer in verbal form to the prevalent and important sort of remark which says 'Be logical; if you think this is wrong you ought also to say that is wrong too'. The re-statement is as follows. Anyone who passes judgment on a particular situation or action has a very wide choice indeed as to what other specific judgments he allows to cohabit with the original one in his moral system; but on the universalisability thesis he must allow some such other judgments, and which other ones he can allow will depend in part upon the nature of the original judgment with which they must cohabit. And it may be that his only way of avoiding allowing his initial judgment to keep unwelcome company is to jerry-build his structure, i.e. to prevent his

original judgment from leading him into a multitude of unwelcome judgments by declaring to be morally relevant some characteristic which he does not really regard, or which he does not wish to be known to regard, as morally relevant.

The upshot of this is that it is not true that once we have settled all the relevant empirical matters we can only agree to disagree on the moral residue, or at best proceed by non-rational techniques of persuasion and the like. It is extraordinary that philosophers should so often have stated this to be the case when it is so patently untrue of what does happen when there is disagreement over moral questions. Strictly moral argument does occur, and sometimes it yields results of one of the sorts which I have infelicitously characterised as sorts of 'embarrassment'. Results of any of these sorts are particularly likely to be reached when someone has made a specific moral judgment which does not seem to accord easily with what he would wish to be taken as his over-all moral position, as for example when a self-declared humanist approves of a hanging, when a Christian approves of a war, or when a democrat condemns the overthrow of a tyrant. But there may also be value of a less dramatic sort in arguments over moral judgments which are not at all obviously misfits. Even a careful and thoughtful person will occasionally allow something which is only a prejudice to masquerade as a moral judgment; and as soon as that happens his system of moral judgments will contain tensions which can be made to cause outright fractures, through the pattern of argument dictated by the universalisability thesis.

Finally: *De gustibus non est disputandum*. Has not this fact provided the main intellectual motivation for opposition to any analysis of moral judgments as not being statements of a special sort of fact? The universalisability thesis provides an important differentia between moral judgments on the one hand and commands, expressions of taste, etc. on

the other; a differentia which precisely and fully answers this particular criticism of 'non-factual' theories of moral judgments by showing how, even on a 'non-factual' theory,

there can be moral argument in a straightforward sense of 'moral' and of 'argument'.