



## Universalizability Without Utilitarianism

Philip Pettit

*Mind*, New Series, Vol. 96, No. 381. (Jan., 1987), pp. 74-82.

Stable URL:

<http://links.jstor.org/sici?sici=0026-4423%28198701%292%3A96%3A381%3C74%3AUWU%3E2.0.CO%3B2-W>

*Mind* is currently published by Oxford University Press.

---

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/oup.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](mailto:support@jstor.org).

## Discussions

### *Universalizability without Utilitarianism*

PHILIP PETTIT

Suppose that an agent—call him *a*—prescribes for himself an action of *F*-ing, which will affect certain other people, *b* and *c*. He prescribes for himself, as we may put it, that *F*(*abc*). *a* prescribes universalizably if and only if he is willing to prescribe that *F*(*xyz*), for any agent *x* and any people affected, *y* and *z*, where *x*, *y*, and *z* are in a situation that is exactly similar in universal properties to the situation of *a*, *b*, and *c*. He is willing to prescribe for appropriate situations that *F*(*bca*), that *F*(*bac*), that *F*(*dfg*), and so on, through other permutations and replacements of individuals.

R. M. Hare maintains that any prescription is universalizable in this sense so long as it is made on the basis of universal features of the case that it governs; he thinks that moral judgement is one such prescription.<sup>1</sup> This claim is compelling. Keep the features the same and you will preserve the reasons. Preserve the reasons and you will have to maintain the prescription. You will have to prescribe, not just that *F*(*abc*), but also that *F*(*xyz*).<sup>2</sup>

Hare's claim has the status of a logical platitude, but he connects it with a further claim to provide a foundation for utilitarianism. This is the claim that 'the requirement to universalize our prescriptions generates utilitarianism'.<sup>3</sup> The upshot is the thesis—the utilitarianism thesis, I shall call it—that any prescription based on universal features of a situation ought to be a utilitarian prescription. The idea is that an informed and consistent agent would be able to maintain such a prescription only if it passed the utilitarian test of maximizing overall preference-satisfaction.

This extraordinary claim is probably the most substantial thesis extant in meta-ethical theory. I wish to contest it by arguing two points: firstly, that the thesis is plausible only for prescriptions based on one very distinctive sort of universal reason,<sup>4</sup> and secondly, that Hare is probably guilty of thinking—as many certainly do—that all reasons are of that type. The first proposition is designed to undermine the utilitarianism thesis, the second to explain Hare's attachment to it.

<sup>1</sup> For Hare's most recent and comprehensive statement of his views see *Moral Thinking*, Oxford University Press, 1981 (henceforth MT). For an excellent overall account of universalizability see W. Rabinowicz, *Universalizability*, Dordrecht, D. Reidel, 1979.

<sup>2</sup> Cf. MT p. 115.

<sup>3</sup> MT p. 111. Hare specifies this claim further, maintaining that it holds in particular for those prescriptions that affect the interests of others. See MT p. 54.

<sup>4</sup> Plausible but not necessarily compelling. See Michael McDermott, 'Hare's Argument for Utilitarianism', *Philosophical Quarterly*, 1983 and Fred Feldman, 'Hare's Proof', *Philosophical Studies*, 1984.

I

(1) *The utilitarianism thesis is plausible only for agent-relative reasons of preference*

Reasons come in two kinds, agent-neutral and agent-relative.<sup>5</sup> An agent-relative reason is one that cannot be fully specified without pronominal back-reference to the person for whom it is a reason. It is the sort of reason provided for an agent by the observation that *he* promised to perform the action in prospect, or that the action is in *his* interest, or that it is to the advantage of *his* children. In each case the motivating consideration involves essential reference to him or his.

An agent-neutral reason is one that can be fully specified without such an indexical device. It is the type of reason which directs the agent to certain properties of the circumstances, the action, the likely consequences of the action, or whatever. Such properties might be: that the situation calls for a radical response, that the action envisaged is the honourable thing to do, that it will have the effect of raising morale, and the like. If regimentation is required, these can all be cast as features of the action, some intrinsic, some relational.

The utilitarianism thesis is asserted of all reasons which involve just universal features of the relevant situation. Universal reasons of this sort may be agent-neutral or agent-relative and we must therefore examine the thesis with regard to both categories. Our question is whether there are any universal reasons, agent-neutral or agent-relative, such that an informed and consistent agent, *a*, can prescribe that *F*(*abc*) on the basis of those reasons, even when the action prescribed does not promise to maximize preference-satisfaction overall.

I shall examine this question first with regard to agent-neutral reasons and then with regard to agent-relative ones. Having maintained that universal reasons of either kind can sustain non-utilitarian prescriptions, I shall then go on to concede that there is one particular sort of agent-relative reasons where this may not be so. I shall show that an informed and consistent agent may have difficulty in prescribing that *F*(*abc*) on the basis of universal agent-relative reasons of preference, if the action prescribed does not promise to maximize preference-satisfaction overall.

*Agent-neutral reasons in general*

Suppose that *a* believes that his *F*-ing promises to instantiate a variety of general properties which, given the alternatives, make it the mandatory option. And suppose that *F*-ing frustrates the preferences of *b* and *c*, the others affected by the action, to the extent that *F*-ing fails to maximize preference-satisfaction all round. *a* might be Socrates, *b* his friends, *c* the others affected, and *F*-ing his taking the hemlock. It seems reasonable to suppose that Socrates thought both that the choice he made was required of him and that it did not maximize preference-satisfaction all round.

<sup>5</sup> On agent-relativity and agent-neutrality see Derek Parfit, 'Prudence, Morality and the Prisoner's Dilemma', *Proceedings of the British Academy*, 1979, pp. 555 ff.; Thomas Nagel, 'The Limits of Objectivity', in S. M. McMurrin, ed., *Tanner Lectures on Human Values*, Vol. 1, Cambridge University Press, 1980, pp. 75-139; and Amartya Sen, 'Rights and Agency', *Philosophy and Public Affairs*, 1982, pp. 3-39. See, too, Philip Pettit and Robert Goodin, 'The Possibility of Special Duties', *Canadian Journal of Philosophy*, forthcoming.

Why should Socrates be guilty of any empirical or logical fault in prescribing, say on the grounds that this is what virtue requires, that  $F(abc)$ ? Why should he not be able to prescribe on just the same basis that  $F(xyz)$ ? The suggestion is that he ought to balk at the prescription that  $F(xaz)$ , as we might put it: that is, at any prescription for a situation where he finds himself in the position of his friends in the actual situation.<sup>6</sup>

The suggestion is unconvincing. Let it be granted that Socrates's friends are desolate and that he recognizes that he, too, would be desolate in the counterpart situations. This observation would not disturb Socrates, for as he believes that his friends are blinded by emotion to the relevant considerations, so will he say that in the situations envisaged he, too, has his vision dimmed by feeling. Why should he allow his present prescription to be affected by a hypothetical prospect in which, due to what is now seen as weakness, he would not endorse the counterpart prescription?

The lesson of this example is obvious. If someone prescribes an action for himself on the basis of some universal agent-neutral reason then, given that he continues to think of the reason as a conclusive consideration, he will unhesitatingly prescribe the corresponding action for counterpart situations. If it turns out that the exercise of universalization gives him pause, then that can only be because he discovers that his original prescription was not produced only by such a reason, but was due in some part to his occupying the agent's role rather than that of anyone else affected. That contingency, he will judge, deluded him into thinking that the agent-neutral consideration was conclusive.

The important point is that there is no intrinsic reason why a non-utilitarian prescription should give way under the influence of such universalization. If the agent bases the prescription on a universal agent-neutral consideration, then he can continue to maintain it over cases where he is the one frustrated. He will do this, so long as he can think that in the hypothetical cases envisaged he is not properly alive to that reason, while in the actual case he is.

### *Agent-relative reasons in general*

So much for prescriptions based on universal agent-neutral reasons. Turning to the other case, we now suppose that the agent,  $a$ , prescribes for himself that  $F(abc)$  on the grounds that  $F$ -ing bears a certain relationship to him or his. Let us say that it enables him to further his career and that it does so at the expense of certain others,  $b$  and  $c$ . Assuming that the reason is universal, the agent is then committed to prescribing that  $F(xyz)$  for any agent  $x$  who is in a position to confer such a benefit on himself; and this, even for those situations where he,  $a$ , is in the position of  $y$  or  $z$ . Can  $a$  endorse that commitment?

As in the case of prescriptions based on agent-neutral reasons, it appears that he certainly can. Why should he balk at prescribing that  $F(abc)$ , just because he recognizes that were he in the position of  $b$  or  $c$  he, like them, would disapprove of the agent's choice? So long as he thinks that the original reason is conclusive, he will explain such disapproval as the product of partial vision. He will continue to think that the reason is conclusive if he places a sufficiently high value on a regime under which individuals strive competitively to promote their careers.

<sup>6</sup> Cf. MT p. 89.

It is clear that there is no relevant difference between prescriptions based on universal agent-relative reasons and prescriptions based on universal agent-neutral ones. If an agent finds such a reason conclusive in his favour, he may equally judge it conclusive when it motivates an action that does him damage. There is no irresistible pressure, such as Hare alleges, that will drive him towards utilitarianism.

*Agent-relative reasons of preference*

Is there nothing, then, to be said for the utilitarianism thesis? Nothing, I believe, in the general case. But there is a special case where the linkage proclaimed in the thesis does have a certain plausibility and I propose now to turn to this.

The case arises when an agent prescribes an action for himself on the grounds, at least in part, that his preferences are thus and so. The reason offered is of the agent-relative sort and, assuming that it is based on universal features of the situation, it commits the agent to prescribing a similar action for any counterpart agent whose preferences assume the same shape. Let us say that he prescribes that  $F(abc)$  simply because his preference is for  $F$ -ing; that he prescribes that  $F(xyz)$  for any agent  $x$  in a similar situation; and that he maintains both prescriptions, despite the fact that the parties affected by the action have stronger preferences for some alternative option available to the agent.

This case may seem to invite treatment along the lines of the earlier example of a prescription based on an agent-relative reason. As the agent in the earlier case was imagined to prescribe that each should promote his own career, even when this fails to maximize preference-satisfaction all round, so the agent here would be seen as prescribing, regardless of utilitarian cost, that each should concern himself with satisfying his own preferences.

The treatment, however, is forced. The agent is imagined in the special case to be concerned that each should satisfy his own preferences but not to care about whether preference-satisfaction overall is maximized, and this unconcern is implicit in his disregarding the utilitarian cost. There is, therefore, a tension in his attitudes which is absent from the general agent-relative case. If I am anxious that people should each promote their own career, I need not worry that under such a regime there is less overall preference-satisfaction than otherwise. If I am concerned that people should each satisfy their own preferences, I may well care that, should they do so, there will be less preference-satisfaction all round.<sup>7</sup>

Consider how I would have to argue in order to defend the view that each ought to satisfy his own preferences, whatever the cost in aggregate preference-satisfaction. I would have to accord supreme importance to my preferences in any situation where I occupy the role of agent, but I would have to discount my preferences for any counterpart situation where I find myself at the receiving end of the agent's action. I would have to put a premium on the fulfilment of my preferences where I, a, can prescribe that  $F(abc)$ , but I would have to be complacent about the frustration of my preferences in the circumstances where I

<sup>7</sup> This care would parallel a natural concern that if agents promote their own careers—the example is scarcely realistic—then career-promotion suffers overall. In each case, as Derek Parfit puts it, we have a principle that is directly collectively self-defeating. See his *Reasons and Persons*, Oxford University Press, 1984, Chap. 1.

occupy the position of b or c. The division of concern and unconcern may not be strictly inconsistent, but it will not come easily.

In this special case, then, I believe that it is plausible to claim that the requirement to universalize one's prescriptions generates utilitarianism. If we imagine an agent whose prescriptions are grounded solely on agent-relative reasons of preference, then we can concede that the requirement to prescribe universalizably might well drive him towards utilitarianism. He would find it hard to defend universal prescriptions whose ground is that the agent satisfies his preferences and whose effect is that, overall, preferences are less fully satisfied than they would be otherwise.

## II

### (2) *Hare probably thinks that all reasons are agent-relative reasons of preference*

If the argument of the last section is sound, then the utilitarianism thesis is false. But the argument plays a second role, for apart from undermining the thesis, it also suggests an explanation for why people like Hare may be attracted to it. This is that they think that all considerations—in particular all universal but non-universalized considerations—which are capable of supporting prescriptions are agent-relative reasons of preference: that is, are considerations which, given the requirement of universalizability, can plausibly support only prescriptions that pass the utilitarian test.

My aim in this section is to provide support for the suggestion that Hare thinks of all reasons in this way. I shall first identify an insensitivity which can easily lead someone to that view of reasons; I do this in order to soften the ground for the suggestion. And then I shall produce two specific pieces of evidence which transform the suggestion into a plausible hypothesis.

The insensitivity which I propose to indict is an insensitivity to a distinction between two sorts of actual or alleged facts: two sorts of considerations, as we may say. They are, on the one side, 'reason-supplying' considerations, and, on the other, 'reason-supposed' ones. The distinction between them comes from the different relation that they bear to the process of practical reasoning.

Every rational prescription can be represented as the conclusion of a piece of practical reasoning, being formulated as 'Let so and so be done!' or 'I must do so and so' or 'So and so is the best thing to do'. The premisses leading to such a conclusion answer to the beliefs and desires that can be invoked to make rational sense of the prescription and, if it occurs, the prescribed action; answering to the belief that p will be the premiss 'p', or 'probably p', to the desire that q something like 'it is desirable that q', and so on.

Reason-supplying considerations are the matters that are explicitly recorded in the premisses of such a practical syllogism. They are the considerations which must be taken to have borne rationally upon the formation of the agent's choice, if the syllogism fairly represents his reasoning. They are assumed to have been relevant to that choice, in so far as they correspond to the beliefs and desires that rationally explain it.<sup>8</sup>

<sup>8</sup> For more on my understanding of action-explanation see Pettit, 'A Priori Principles and Action-Explanation', *Analysis*, 1986 and 'Broad-minded Explanation and Psychology' in Philip Pettit and John McDowell, eds., *Subject, Thought and Context*, Oxford University Press, 1986.

Reason-supposed considerations bear on these latter attitudinal items. They are the facts which, while they are not themselves mentioned in the premisses of the syllogism, must have obtained if those premisses are relevant. If the premiss 'p' is relevant, then it must be that the agent believed that p; if the premiss 'it is desirable that q', it must be that he found it desirable that q or—I take this as a paraphrase—that he desired that q.

Suppose that someone rationally prescribes for himself that he should help an old lady across the street. His reasoning may be susceptible to one of the following representations: 'The old lady needs help; therefore I must offer it', or 'It would be kind to lead the old lady across the street; therefore I must do so', or 'It is good to help the elderly; this is an elderly person in need; therefore I must help her', or 'A considerate person would lead the old lady across the street; I want to act considerately; therefore I must lead her across'.

If the syllogism is a fair representation of his reasoning, then the facts alleged in the premisses supply the agent with the reasons that affect his choice. However, each such fact supplies a reason only so far as some corresponding fact is supposed about the reasoner. If 'the old lady needs help' is relevant, he must believe that that is so and he must want to provide help in such a case; if 'I want to act considerately' is relevant, he must again have this belief about himself and he must want to satisfy the preference in question.

It should be noted that this claim should be conceded, not just by non-cognitivists, but by those who think that some considerations are intrinsically reason-giving. A consideration may be a reason for doing something independently of whether the agent has a corresponding desire, and its motivating power may be needed to explain the presence of that desire, therefore, rather than the other way round. Nevertheless, the agent must indeed have the desire if the consideration deserves mention in a practical syllogism that purports to represent his reasoning. The consideration cannot be a reason that moves him if he lacks the desire that corresponds to it.

When an agent prescribes an action for himself, it is important to know what the reason-supplying, and what therefore the reason-supposed, considerations are. Did he act on the grounds that R, given that he had the appropriate R-belief and R-desire? Did he assist the old lady on the grounds that she needed help, given that he believed that she was in such need and that he desired to help? Or did he act on the grounds that he had the R-belief and the R-desire, given that he had the background attitudes appropriate to such reasons: the meta-R-belief (i.e., that he had the R-belief and the R-desire) and the meta-R-desire (i.e., to satisfy the R-desire)? Or was his a mixed case? Did he act, for example, on the grounds that R and that he had the R-desire, given that he believed these things and had the meta-R-desire? Did he act on the grounds that the old lady needed help and that he desired to offer assistance, given that he believed these things and had the meta-desire to satisfy his desire to assist?

I believe that an insensitivity to the distinction between reason-supplying and reason-supposed considerations is widespread in contemporary philosophy and that it is responsible for a variety of mistakes in moral and related thinking.<sup>9</sup> My

<sup>9</sup> The insensitivity amounts to a failure to distinguish between the context of deliberation and the context of explanation in considering reasons. I have traced some of its other effects in an unpublished piece on 'The Normative Role of Decision Theory'. One further effect, as Simon Blackburn has

interest in it here stems from the fact that it serves to generate the assumption that all reasons are agent-relative reasons of preference.

The transition to that assumption comes with the thought that whenever a consideration R is presented as an agent's reason for prescribing something, and R does not itself involve an appropriate R-desire, then the fully specified reason is the combined consideration that R and that the agent has the R-desire, a thought that is seductive if one does not distinguish clearly between reason-supplying and reason-supposed considerations. For any R, as we have already seen, the agent's R-desire must have a place among the facts supposed by that reason. Thus, it may easily be imagined that the same holds for the facts supplying the reason.

The upshot of our discussion, then, is that it is extremely easy to slip into the view that all reasons for prescribing something are agent-relative reasons of preference. At this point we turn to consider two pieces of evidence which serve, against that background, to motivate an ascription of the view to Hare.

The first piece of evidence is that he is much given to a weighing metaphor in characterizing practical reasoning and that he interprets this metaphor in a way which insinuates the assumption that an agent's reason for prescribing anything must mention his desires. The metaphor represents practical reasoning as a competition between conflicting desires and suggests that the desire which is the heaviest or strongest—sometimes the conflictual aspect of the metaphor comes to the fore—is that which necessarily prevails. 'Our other preferences may outweigh this one; what we prefer all in all is determined by the balance of them without external constraint.'<sup>10</sup>

Under one interpretation, this metaphor is to be understood mechanically; under the other, it is to be taken metricaly. The mechanical idea, neither subtle nor insidious, is that deliberative conflict is like a physical disequilibrium whose resolution depends, at least in part, on which of a number of elements weighs most. As the spinning die tends to come down on its heaviest side, so the undecided agent is swayed by his weightiest or strongest desire.

The metrical interpretation of the metaphor is rather different. Here, deliberative conflict is pictured on the model of a competition to determine the heaviest of a set of objects by weighing them in a balance. Desires each get their turn on the scales and victory goes to the one which notches the highest mark. It outweighs all competitors.

The mechanical model is an everyday analogue of deliberation and does not beg any particular questions. The metrical model is more contentious. It is easily introduced once we have begun to think of desires as having weights, but it begs a serious issue by imposing the assumption that the resolution of desire-conflict requires an umpire. There has to be someone who sees which desire outweighs competitors; there has to be someone, indeed, to weigh the desires against one another.

The assumption that an umpire is necessary means that we have to picture the undecided agent as calling up each of his conflicting desires for measurement and

suggested to me, may be to motivate too facile a criticism of certain non-cognitivist theories: the criticism would be that in stressing the place of desire, those theories overlook the fact that many of our reasons for action do not involve our desires.

<sup>10</sup> MT p. 225.



comparison. Such an agent will eventually prescribe actions on the grounds that R, where this is the content of his relevant beliefs, and that he has an R-desire which prevails over all opposition. The old lady needs help and his desire to provide assistance triumphs over any conflicting preferences.

The weighing metaphor pervades Hare's writing and is clearly subject to the metrical interpretation. He refers to weighing preferences in a pair of scales, speaks of one or another tilting the balance, and talks of how they compare in weight or strength.<sup>11</sup> Sometimes the model dominates whole passages, as in the following example.

We shall in any case do what the balance of our present preferences requires—in other words, act on the prescription which results when the prescriptions we are now disposed to accept are balanced against one another in proportion to the strengths with which we accept them . . . And in this process we shall (since it is now that we are making the decision) rationally consider only our now-for-now and now-for-then preferences.<sup>12</sup>

The presence of this metrical model in Hare's writing is powerful evidence of his assuming that an agent's reason for prescribing something must always mention his desires. Without that assumption, it is hard to see how he could find the model natural or even tolerable.

The second piece of evidence that Hare makes this assumption comes from his way of characterizing prescription in relation to preference. He says that 'to have a preference is to accept a prescription' and that 'all prescriptions are expressions of preferences or of desires in a wide sense'.<sup>13</sup> This means that an agent always prescribes what he most prefers or, alternatively, that he always prescribes that his preferences are satisfied. These remarks are capable of innocent construal, but Hare takes them in a manner that intrudes the assumption that an agent's reason for prescribing something always mentions his preferences or desires.

Innocently interpreted, the claim that someone prescribes what he most prefers can be understood as follows: for any option such that he most prefers it, the agent prescribes that option. Similarly, the remark that the agent prescribes that his preferences are satisfied can be read in this way: for any option such that it satisfies his preferences, the agent prescribes that option. Under interpretations of this kind, these modes of speech beg no questions about the form of the reasons which move the agent. He may be moved by the desirability of the option in question, so that the reason does not refer to his preferences, or he may indeed be moved by the consideration that he prefers that option.

Under the alternative interpretations, the remarks lose this innocence. They intrude the assumption that, in prescribing anything, an agent always reflects explicitly on the state of his preferences. The interpretation takes the phrases in which the prescribed object is described as representing in each case not just the object of the prescription, but the agent's way of thinking of that object. The agent is supposed to think of what he prescribes as 'what he most prefers' or as 'that his preferences be satisfied'. If he does think of the object of his prescription in this way of course, then his reason for prescribing it cannot fail to include a reference to the state of his desires. It cannot fail to be an agent-relative reason of preference.

<sup>11</sup> See, for example, MT pp. 110, 112, 125, 128–9, 180–1.

<sup>12</sup> MT p. 104.

<sup>13</sup> MT pp. 91 and 185.

It is clear from many passages in his writing that Hare adopts this alternative style of interpretation when he characterizes the objects of prescription by reference to what the agent prefers. These two passages may make the point.

To recognise that that person would be myself is already to be prescribing that, other things being equal, the preferences and prescriptions of that person should be satisfied.<sup>14</sup>

If I were going forthwith to have the preferences which he actually has, I must now prescribe that they should be satisfied.<sup>15</sup>

The fact that Hare characterizes the objects of prescription by reference to preference, and that he interprets the characterization in the question-begging way, serves as a second piece of evidence that he makes the assumption I have sought to pin on him. He takes it that an agent's reason for prescribing something must always mention his desires or preferences.

This, then, is the denouement. We saw in section I that the utilitarianism thesis is plausible only for prescriptions based on agent-relative reasons of preference. What we have seen in this is that there is good, if more or less circumstantial, evidence that Hare takes all reasons to be of that kind. The upshot is not unsatisfactory, for even if this evidence had been unavailable, we would have been tempted to ascribe such a view of reasons to Hare. How otherwise could we have explained his attachment to the extraordinary claim that all prescriptions based on universal reasons ought to pass the utilitarian test?<sup>16</sup>

*Research School of Social Sciences  
Australian National University  
Canberra, A.C.T. 2601*

PHILIP PETTIT

<sup>14</sup> MT p. 221.

<sup>15</sup> MT p. 223.

<sup>16</sup> I am indebted, for comments received on an earlier draft, to Simon Blackburn, Bob Goodin, Karen Jones, Christie Slade, Jack Smart, and Albert Weale.