CHAPTER IX

Foundations

We come now to the sometime vexed question of whether morality, and the utilitarian theory in particular, can be supported by proof. The word “proof” must no doubt be used advisedly in this connection, but it may be that these scruples should trouble us less than they have. At any rate, I do propose to offer what seems to me a rational argument for adopting the principle of utility, and against adopting any other “conception” of ethics.

What Are We To Prove?

Some philosophers have held, on occasion, that morality not only needs but can have no “proof,” no justification. They reason thus: To prove that I ought to do my duty is to endeavor to show that I ought to do what I ought to do. But this is nonsense. If I ought to do something, then there can be no additional proof, and certainly need be no additional proof, that I ought to do what is so characterized (namely, as what I ought to do). Either I can offer a worse reason, a lame one, such as that I might enjoy it—but then, what if, as is often the case, I would not?—or I can offer the same old “reason” over again, viz., that it is my duty. Nothing stronger is available.¹

This argument is decisive against its particular target. To call an act right or a duty is to say that one ought to do it, and so, to ask for “proof” for such statements as “you ought to do what

¹This, of course, is taken substantially from H. A. Prichard’s “Does Moral Philosophy Rest on a Mistake?” in Moral Obligation (New York: Oxford University Press, Inc., 1949).
MORALITY AND UTILITY

is your duty” is simply misguided. But this is not the problem. The man who is worried about morality will not, in all likelihood, have his worry removed by such an argument. Just as King George wanted to know whether Scott was the author of “Waverley,” and not whether Scott was Scott, so our man wants to know, if we must put the above argument to him, not whether we ought to do our duty, but whether anything is our duty, and in particular how we can know this.

In order to be able to offer a general, single-barrelled “proof” of morality, one must have a general, single-barrelled characterization of it which is nevertheless nontrivial. If we conceive of morality as consisting of a number of unrelated principles, then if one is to supply proofs, one will have to supply a different proof for each principle. Fortunately, utilitarianism is par excellence a single-barrelled theory, in this sense. It proposes that all right acts are utility-promoting, all wrong ones utility-defeating—in other words, that utility is coextensive with moral value. This being a general, but substantive, thesis, it makes sense to try to prove it. This ease of formulation is probably why the question of proof arises so naturally in the context of utilitarianism. It would tend to come up in any “monistic” theory of ethics clear enough to be discussable. (Many are not: that one’s fundamental duty is to “realize the self” or to harmonize with the Absolute, for example, are contentions which inhibit sensible discussion.)

What Constitutes Proof?

When do we have a proof and when do we not? It is hard to give an informative answer to this. Doubtless, a good argument is one that is “capable of moving the mind” in the general direction of the conclusion: but what does this come to?

At least we can say what it is not. The skeptic, about whom more will be said below, very often seems to be demanding that we literally talk people into being good. This can sometimes be done, and is certainly a worthy aim; but to set it up as a general standard for validity would be quite unreasonable. For one thing, even people who believe that x is what they ought to do
sometimes do not do it. It is even open to question whether a person who believes he ought to do x necessarily wants to do x. One cannot deny that there is some connection between the acceptance of moral principles and acting as they prescribe, but this seems too much. At very most, we must aim at convincing people that they ought to act in conformance with the principle of utility. But this isn't quite right either, for if our arguments in Chapter I are sound, then most of our readers already do believe this, since they believe that they ought to keep their promises, relieve suffering, avoid harming people, and so forth. What we apparently need to do is to produce a feeling of "rational conviction" in them. What is this?

To begin with, it is unfair to call rational conviction a "feeling," just like that. To have such a "feeling" is to think (that is, to be rather confident) that one knows whatever the conviction is about. Under what circumstances is one entitled to do this—particularly in the case of ethical matters? I do not think that we need a special characterization of this for ethical questions as distinct from others: to know is to have good reason for believing. And to have good reason for believing is to see that certain promises are true, and that they imply the conclusion in question.

It might be held that knowing is not always reasoned in the suggested sense. When I spot the bluebird in my back yard, I know that there is a bluebird in my back yard. Where is the "argument" here? Where the premisses and the conclusion? But this overlooks the distinction between knowledge and those perceptual states that (in the relevant cases) are its prerequisites. Seeing is perceiving, but it is not believing: it is merely the prerequisite to believing, when what you believe is about what can be seen. And knowing is being able to produce a satisfactory defense of what you believe. "There was a bluebird in the back-yard this morning." "How do you know?" "I saw it." (If this continues, "How do you know that you saw it?", the answer might be, "Well, it was a perfectly clear morning and the branch is just six feet from our kitchen window," or any number of other things. But satisfaction is obtained when the knower shows that his utterance has been properly cued by his experience, and this can always involve an argument.)
MORALITY AND UTILITY

Similarly, knowing that such-and-such is a duty is being able to defend satisfactorily the proposition that it is a duty. Such a defense consists in supplying premises which all parties to the argument accept, and then showing that the conclusion ("It was my duty," or whatever) follows. Of course, the relevant premises might themselves be called into question by some philosophical questioner, and then a new argument may be needed. When does this process end? It ends when we show the questioner that the denial of what we are out to prove would involve an absurdity or contradiction.

Some claim that not all absurdities involve contradiction. This may be so, but they are all near to it in one respect: namely, that we convict a man of absurdity when we find him maintaining (verbally) a view which contradicts either some of his other statements, or some statements which his behavior shows that he accepts. And when does one's behavior "show" that one accepts a statement, especially when one purports to be denying it verbally? When the behavior is of the kind which it is the purpose of the sentences in question to signify. If a man hits his finger with a hammer, he says "Ouch!"; and thus we know it hurt. If he claims that it didn't, and the exclamation was loud enough and clear enough, we conclude that he's kidding, or acting, or insincere, or that he hasn't mastered the meaning of the word 'hurt.'

It is in this way that we want to know what is right and wrong, and proof consists in getting us into this state.

Is Proof of Ethical Statements Needed?

In the Introduction, I gave reasons for thinking that proof in ethical matters was not impossible. The question might remain whether it is necessary, which has often been denied. Those who deny it, however, tend to maintain that "proof is as needless as it is impossible"; if they thought they could supply it, they would not deny that it was necessary.

The words 'know' and 'prove' are intimately related. The man who knows, as I suggested above, is the man who can give
a good argument. And arguments are needed when what we believe is called into question. Whether proof of ethical statements is needed, then, depends on whether they are called into question. But of course, they often are, both in particular and in general. I contend, moreover, that there is no difference in principle between the type of argument called for in the case of the "general" question, and that called for in the "particular" one, though of course there will be differences in generality. The philosopher's business is to show with precision ("as much as the subject permits," no doubt; but this may be quite a bit) what the rock-bottom steps in such arguments should or could be.

Moreover, as Baier points out, morality is an institution peculiarly requiring rational support. For it frequently requires us to do what we wouldn't otherwise have wanted to do; it calls for sacrifice of one degree or another, and sometimes these are serious. Such sacrifices need justifying, and it is the rational person in particular to whom the need for such justification will occur. Others may be "conned," duped, brainwashed, or otherwise habituated to unquestioning obedience, but the reflective will raise doubts. If these cannot be laid to rest, then suspicion arises that the whole institution may be a fraud. In any case, it could lose the support of these individuals, and this may well lead to a chain-reaction, as others come to ask why they too should not participate in the "luxury" of immorality.

Method

How do we, then, construct an argument both powerful enough and general enough to "support" so weighty an institution as morality? If my account of knowing is close enough to do, the general outline becomes clear enough. We need to search about for premises which anyone, or at least anyone capable both of participating in and raising questions about morality, will accept. But what sort of premises might these be?

MORALITY AND UTILITY

Clearly, the best kind of premises would be those we can with a clear conscience think of as “necessary” or a priori. More precisely, we want to search for premises which it would make no sense for a person who qualified as a member of our audience (which presumably is everyone, but possible exceptions will be noted eventually) to deny. Among these, certainly, will be principles of formal logic, but of course these will not be sufficient. We shall also need, most especially, analyses of the relevant portions of evaluative language. If we can show that the “very meaning” of expressions which our audience must employ is such that to deny the principle of utility would land one in absurdity, then the proof will be complete. This calls, then, for an excursion into meta-ethics, which we have been to some extent eschewing in former chapters but clearly can no longer ignore.

False Starts

There are a number of pitfalls to be avoided, which happily the labors of recent theorists have pointed out for all to see. For example, there are the various forms of the definist fallacy, and the “naturalistic” fallacy in some understandings thereof. In Chapter II we have hopefully warded off such spectres; that was the point of denying that the principle of utility is “analytic.” But the fallaciousness of such fallacies is exposed by the conditions of proof lately laid down. To adopt a definition that doesn’t define the expressions in which the other chap’s questions are couched is to lose your hold on him for argumentative purposes. If Jones doesn’t mean by ‘good’ what my proposed definition says ‘good’ means, then the fact that utility-production is “good” in that sense is simply uninteresting. This, I take it, is a fundamental impetus behind “ordinary language” philosophy.

The same general consideration defeats many other possible efforts. Any sort of intuitionism, for example, fails as soon as

---

3 Why this does not (hopefully) jeopardize the views in Ch. II about the “synthetic” status of utilitarianism will come out in the sections below. Cf., especially “Proof of Utilitarianism,” and “Mill’s Proof.”
the audience fails to share our supposed "intuitions."¹⁴ Appeal to theological premises is of no use, both because numerous members of our audience will not share them, and for various logical reasons well understood by philosophers.⁵ Appeal to "human nature" in any very substantial sense is dangerous because implausible.⁶ There is little point in saying that all men have the natural desire to help others, because the facts show that either they don't, or it's too weak to be of service in numerous cases.

The line that everyone would like to be able to take, and which some have attempted to take, is that of showing that in some way or other, morality is in one's interest. But the cruder forms of this pretty obviously won't work. One might be interested in making a lot of money, owning a yacht, keeping a mistress, and so on; and it is all too easy to conceive of situations in which cheating, breaking promises, or killing a few enemies will in fact help one to achieve these goals. The wily prudentialist will retort that these interests are of a kind the satisfaction of which won't really make their possessor happy. And he might even be right. Unfortunately, there is a good deal of evidence to show that he's wrong, if 'happiness' is not stretched in some way; but if it is stretched, then we have the same problem as the various fallacious starts mentioned above come up against. Our audience may not be interested in being "happy" in the newly defined sense.

It may be that in some extremely subtle sense, morality is in one's interest. But we must be careful of various obscure logical traps. For example, it is indeed true that if everyone were to promote [note: not 'try to promote,' but 'to promote'] everyone's happiness in the sense defined in Chapter II above, this would be in everyone's interest. Such is the premise of Baier's attractive and ingenious argument.⁷ If everyone follows the rules of morality (rules which occasionally run counter to self-interest

---


⁵ See n. 26, Ch. VIII.


⁷ Baier, *Moral Point of View*, Ch. 12.
for everyone), then everyone would be better off. Therefore, I should follow the rules. . . . If . . . For as Baier recognizes, if I can depend on no one else to follow them, it is contrary to my interest to follow them, and so only if I have reason to suppose others will follow them, have I reason to follow them, and so should follow them. So everyone has reason to be moral so long as he is reasonably sure that others will be, too.

This argument requires close consideration; it is surely somewhere near the truth. But there are certain difficulties which may be decisive, and at least give us reason to hope we can do better. In the first place, we must be careful how we are to take the argument. One possible interpretation would be that “the best average life is a sufficient recommendation of the moral point of view, that the benefit of the group justifies acceptance by each individual,” as David B. Gauthier puts it. The trouble here is that the argument seems unintuitive. We seem to be able to ask, “But why does the benefit of the group justify acceptance by me, if I can gain from a more selfish policy?” There is no use in saying that it just does, if our hope is to convince all rational beings; for it is precisely some rational beings (such as myself and, as it turns out, Baier) who question it. The other interpretation is that the argument shows that an individual can expect more satisfaction from adopting morality than from not doing so. The trouble is that even though this is perfectly true at the outset, yet there will come a time when it no longer is, namely when, as Baier admits and insists will happen from time to time, morality requires a genuine sacrifice of me. At this point, it ceases to be true that the policy offers expectation of maximum benefit, even though at the outset I had reason to suppose it would. So I no longer have a self-interested reason to continue to follow the moral “policy.” If self-interest really is the foundation of the policy, then it breaks down here. Self-interest, then, cannot sanction the general adoption of anything but a policy of (genuinely enlightened, to be sure) self-interest.

Finally, Baier’s argument seems to work only among people whose power over others is limited. An extremely powerful individual who could control others as he wished, would appar-

---
9 Baier, *Moral Point of View*, p. 314: “But while enlightened self-interest does not require any genuine sacrifice from anyone, morality does.”
ently not have a reason to be moral by this argument (taking it, of course, in the second sense). This seems unsatisfactory, for surely powerful people ought to help people, or at least not to harm them, just as persons of ordinary power ought to. It is absurd to make duties dependent on power in this particular way, Nietszche to the contrary notwithstanding.

What we need, in short, is an argument, universally acceptable, which makes morality “coordinate with prudence,” as Gauthier puts it. The fact that others will suffer must be shown to be just as good a reason for avoiding it as the fact that I will suffer. No argument which is based essentially on considerations of self-interest, it appears, can do this. What does this leave?

Self-interest

The form of our question has gradually evolved into one of self-interest versus the general interest. The reader will, I think, have considered this a natural evolution. It is unsurprising to be told that being moral is being unselfish. What makes it unsurprising is that nobody doubts that the fact that x would be in my interest is a good reason for doing x. The work of recent ethical theorists, as well as many past ones, suggests that

10 Gauthier, p. 109, accuses Baier of this result. It is difficult to interpret the sense of the argument in Moral Point of View on this question, but in a symposium at the University of Waterloo (April 23, 1965), Baier agreed in a question period that so far as he could see, it would not be possible to have moral relations with a supremely powerful person (God, for example).

11 According to whom: “One has duties only to one’s equals.” (Beyond Good and Evil, Ch. IX.) This selection is found in Philosophic Problems, ed. Mandelbaum, Gramlich, and Anderson (New York: Macmillan Co., 1957), p. 370.

12 E.g., Gauthier, Sparshott, Baier, Nowell-Smith, and Castañeda in “Imperatives, Decisions, and ‘Oughts’” (Morality and the Language of Conduct, pp. 219–92), and probably Toulmin, Hare, and Stevenson, among others. All of them say or imply that if one wants something, getting it will not interfere with the satisfactions of others, and it isn’t contrary to any other important wants of one’s own, then it would make no real sense to deny that I am justified in trying to get it.

13 Plato, Aristotle, all of the Stoics and Epicureans in one way or another, and probably most of the medievals, would accept this; all, however, have special views as to what really is in one’s interest, which makes it a bit more doubtful. Kant, incidentally, implies this in his famous view that hypothetical imperatives are “analytic.”
there is something a priori about this, if I may put it so. It is felt that there is a logical connection between wanting something and its being good, or at least between wanting it and calling it good. (Hobbes: "whatsoever is the object of any man's appetite or desire, that he for his part calleth good.") The logical niceties of this connection will not be explored very thoroughly here, though part of the reason of it will be suggested below. But at any rate, we may sum this up by saying that each person assumes that the fact that something is in his interest makes it good, prima facie. 'Prima facie' needs to be added for two reasons. First, interests or wants sometimes conflict, and then one will have to adjudge one less good than the other, to the point where sometimes one will adjudge a want downright bad, as interfering with one's better wishes. And second, it is not self-contradictory to perform an action which is seen to be contrary to one's interest as a whole (unless we trivialize it by calling one's moral principles one's "interests").

Again it is necessary to warn against misinterpretations of this. Nothing about "human nature" is being assumed here, and nothing about the internal character of one's interests. These may be to any degree ascetic, inspiring, or what-have-you; they may also to any degree be essentially directed at the welfare of others, though as a matter of common observation they often are not. In order to understand the status of this principle, and especially its bearing on the construction of our argument, we must next consider the logic of evaluative expressions.

The Logic of Evaluation

What (if anything) are we saying about something when we call it good? By this time, philosophical analysis has reduced the area of reasonable disagreement to a rather narrow circle. It is agreed on all hands that to say that something is good is not,

---

15 The accounts in Nowell-Smith, Sparshott, Baier, and Gauthier all deserve careful study; I would not presume to improve on them in this matter, nor to be able to do so here.

266
in the most obvious sense of the word "quality," to attribute a
quality to it.\(^{16}\) It is, at least and in some way or other, to come
out in favor of it, to recommend it. On the other hand, though
goodness is not a quality, the man who calls something good is
necessarily attributing some or other qualities (in the ordinary
sense of "quality") to it. In calling something good, one is say-
ing that there are reasons for preferring it, for being in favor of
it, for using it, or whatever the appropriate attitude or activity
may be. To say that there is a reason for preferring this par-
ticular thing is to say that it is of such a kind as to be preferred:
to say, in short, that it is to be preferred because it is the sort
of thing it is, or has the qualities it does.

Saying that something is good is not just expressing approval,
then; it is purporting to have reasons for approval. It may or
may not be that we could say, instead of "this is good," "this
has qualities A, B, C, . . . so, take this!," or again, "this has qual-
ities A, B, C, hooray!"\(^{17}\) Quite likely each is correct in some
contexts, incorrect in others. But the essential point is that in
evaluating, we are supporting choices, preferences, lines of action
which are alternative to other possible ones. We are, one can
equally well say, defending them. No wonder, then, that we
think of our wants as\(^ {16}\)\(^ {17}\) *prima facie* conferring value on their ob-
jects. For if I am willing to pursue a want, I have a reason for
defending it. The man who condemns a certain line of activity
in himself, while doing it, condemns himself, in a sense; and if
he applies the language appropriate to condemnation but seems
totally cheerful about it, we may assume that he is insincere.
Applying evaluative language puts logical pressure behind some
choice or activity.

In short, evaluating is reasoning: supplying reasons in support
of choices and the like. From this, it follows logically that the
application of evaluative predicates is intrinsically "generalizing."

\(^{16}\) Sparshott, *An Enquiry into Goodness*, Ch. 6, finds reason to call good-
ness a "quality," but also reason not to. His subtle discussion of this is
worth close attention. But it is, I think, compatible with the account
attributed to "philosophical analysis" here.

\(^{17}\) This, of course, is from Stevenson's "second pattern of analysis." Cf.,
*Ethics and Language* (New Haven, Conn.: Yale University Press, 1944),
Ch. IX. See also the useful criticism of George Kerner's *The Revolution
MORALITY AND UTILITY

To say of a thing that it is good is to voice a preference or to suggest or advocate preferring it, because it is A, B, C . . . . But to say this is to say that if another item is also A, B, C, then it too is to be preferred. I can’t say both that x is good because it is A and that y is bad (or indifferent) because it is A, unless I have further explanations.18

These two features of evaluative language are, I think it fair to say, pretty universally regarded as essential to it; the same goes for the “language of obligation,” “ought,” “wrong,” and so on. The difference, as we have suggested in Chapter III, is mainly that “obligational” language is stronger in its committing functions. These features are the ones concentrated upon by Hare,19 whose argument is substantially similar to the one I shall develop.

Refutation of Egoism.

G. E. Moore was the first to produce a refutation, on purely logical grounds, of egoism, the ethical view that everyone ought to act only on reasons of self-interest.20 He has since been followed by others whose treatments differ in detail from Moore’s.21 I shall follow a similar line of reasoning.

Egoism is a theory, not an attitude. Let us, for the sake of simplicity, call the latter ‘egotism,’ denoting by this the tendency to act in the way recommended by egoists. The egoist, then, is the philosopher who recommends to all men that they pursue only their own interests, taking no essential account of anyone else’s. Note that he is a theorist: he is arguing that this policy has the support of reason. This is equivalent, I think, to saying that the egoist says that we ought to follow self-interest only, that this is the best thing for everyone to do. To regard egoism as an ethical theory is to regard it as applying the moral vocabulary systematically on the basis of self-interest.

18 See Kerner, pp. 21–24, especially 32 ff.
19 See Hare, Freedom and Reason, Part I.
FOUNDATIONS

Consider, then, the situation in which two people, A and B, have incompatible interests, in the sense that A's is satisfied if and only if B's is dissatisfied. Clearly there are frequent occasions on which this happens. Since the egoist holds that the satisfaction of A's interest is right, he is obliged by logic to hold that the dissatisfaction of B's is right; but since he is also committed to holding that the satisfaction of B's interest is right (hence that the satisfaction of A's is wrong), we have a contradiction. The satisfaction of either interest is both right and wrong. The egoist is rooting for both teams; which would be rather odd, but not self-contradictory, were it not that he claims to be rooting with the support of reason. It is this claim that proves fatal to his theory.

We must be careful to ward off a misconception about this argument. Some would reply that after all, the egoist is merely saying that if doing x would satisfy A at the expense of B, A would be justified in doing x “from his point of view,” and that if doing y would satisfy B at the expense of A, B would be justified in doing y from his point of view. But this is an evasion, as G. E. Moore correctly saw in discussing the expression “my good”:

In what sense can a thing be good for me? It is obvious, if we reflect, that the only thing which can belong to me, which can be mine, is something which is good, and not the fact that it is good. When, therefore, I talk of anything I get as ‘my own good,’ I must mean either that the thing I get is good, or that my possessing it is good. In both cases it is only the thing or the possession of it which is mine, and not the goodness of that thing or that possession. There is no longer any meaning in attaching the ‘my’ to our predicate and saying: The possession of this by me is my good. Even if we interpret this by ‘My possession of this is what I think good’ the same still holds: for what I think is that my possession of it is good simply; and, if I think rightly, then the truth is that my possession of it is good simply—not in any sense, my good; and if I think wrongly, it is not good at all . . .

Some have missed the point Moore is making here, by supposing that it depends, for its acceptability, on his theory that ‘good’

22 Moore, Principia Ethica, pp. 98–99.
MORALITY AND UTILITY

connotes a special property. This is not so. The word 'good,' and other evaluative words, are logically incapable of denoting any sort of “private” objects (a fact, incidentally, which may itself be employed to refute the theory that goodness is a simple intuitable property). As Moore saw, to say that so-and-so is my good is either to say that I think it good, or that its being mine is good. To say that I think it is my good could mean that I think that its being mine is good; but in this last sentence, it is nonsensical to qualify ‘good’ by ‘my’ or by ‘according to me.’ In saying that I think it is good, all the qualifying that needs to be done is done. “I think that I think it’s good” adds nothing to, and perhaps subtracts something (viz., intelligibility) from ‘I think that it’s good.’ And if one supposes that ‘according to me,’ or ‘I think...’ must always be added to any occurrence of ‘good,’ infinite regress looms large before us: for even in ‘I think that I think it’s good,’ we have an occurrence of ‘good,’ so that the whole would again have to be expanded into ‘I think that I think that I think...’, which is as ridiculous as it is boring.

If the egoist were saying nothing more than that when something is in a person’s interest, he, for his part, will call it good, then this is not only perfectly acceptable but has already been insisted upon above. Only it isn’t ethical theory. An ethical theory (as distinct from a meta-ethical theory) must use evaluative expressions, not mention them. Theories about what people think is right or wrong are not theories about what is right or wrong. In order to become such, a theory must side with some or other of the parties in question, or declare both mistaken. The trouble with egoism is simply that it attempts to side with all parties at once, which is contradictory when some of them disagree with others.

Another misinterpretation of the foregoing refutation of egoism goes as follows: What the egoist is saying is not that A ought to do x and B ought to do y, but rather that A ought to try to do x and B ought to try to do y. What about this? It is, in fact, either another evasion, or else a shift in the theory. For suppose we were to ask, of the conflict which would result if A tries to do x and B to do y where x and y cannot both be done, Which party ought to win? The egoist must either answer or refuse to answer this question. If he answers that one or the other ought
to, then he is immediately contradicted by his principle that anybody's interests justify him in doing what would satisfy them, and the refutation is the same as before. He might refuse to answer, or he might answer, "may the best man win!" These, however, come to the same thing. If the egoist claims that his theory only issues in sanctions for trying, and not for succeeding, at any given aim, then he is tacitly underwriting the stronger side. "May the best man win!" would contradict egoism if the application of 'best' were determined on any other basis than that of satisfaction of the interest of the party concerned. But the trouble is that on that basis, we get contradictions, since the interests are incompatible. And our question, remember, was "who ought to win in this contest (quarrel, fight)?" As a reply to this question, the answer "I refuse to answer," or the answer "the stronger" are both equivalent to, "Well, whoever wins ought to win." This is a new theory altogether: the theory, namely, that whoever does win in contests due to conflicting interests ought to win. But this is incompatible with egoism. For consider little guys who can't win, or just anybody who doesn't like to fight. The theory that those who would win at fights ought to win conflicts with their interests, and hence with a theory that says that whatever your interests are, they ought to be achieved. The doctrine that might makes right is a different doctrine from egoism. Pure egoism, then, is inextricably ensnarled in contradiction.

Proof of Utilitarianism

Let us consider what really results when we take seriously the evaluative principle assumed by each person, namely that "the satisfaction of my interests prima facie is good (hence, prima facie justifies me in doing whatever will attain it)." We have seen that if we generalize this theory leaving off the 'prima facie,' we get contradictions. But why, it might be asked, do we need to generalize at all? Let us, by way of reply, consider a "non-generalized egoism" and see what happens.

Suppose, then, that someone promulgates the theory that the
criterion of right and wrong is satisfaction, or dissatisfaction, of Jones' interests, where Jones is some particular person. We do not, at the moment, much care who he is, and it is not even necessary that the advocate be Jones himself: Jones might be a supposed Messiah, and the advocate one of his prophets. (In fact, certain religions approximate rather closely to just this model: the ultimate criterion of right and wrong is satisfaction of God's interests, and this view is advocated by priests, prophets, and miscellaneous clerics.) What is important is merely that we remember that it is satisfaction of his interests that is made the ultimate criterion, to which no further appeal is supposed to be possible.

Now let us ask the question, why ought we to accept this "principle" that satisfaction of Jones' interests is the ultimate criterion of right and wrong? It might be replied that since it is being put forward as an ultimate criterion, no answer to this question could possibly be legitimate. (If I read him rightly, this is precisely the view of Kierkegaard, and probably most "existentialist" theologians, about God). This reply is a misunderstanding. The fact that something is an ultimate criterion can always be explained, though of course the criterion itself cannot be deduced from a further criterion. But the latter is not what we are asking; only the former.

This question, Why should we accept Jones' interests as the ultimate criterion of right and wrong?, is equivalent to asking why we should accept Jones' interests, not why we should accept Jones' interests for this criterion. For we would all like our own interests to be such a criterion, and therefore can understand perfectly well what the word 'interests' is doing there. What we want to know is, why Jones'?

There can be only two answers to this question. The first possibility is "Because Jones' interests are of type ϕ"; the second is, "Because Jones is Jones, and that's the end of it." Let us examine each by turn.

---

23 One gets this impression from Fear and Trembling (New York: Anchor Books, Doubleday and Co., Inc., 1954); but Professor Alistaire MacKinnon assures me that it does not apply to the later Kierkegaard. On the general issue, cf., my "Existence and Particularity," Southern Journal of Philosophy (Spring, 1965), and further discussion below.
The first reply will not do at all, for it contradicts the supposition that Jones' interests are the ultimate criterion. Perhaps all of Jones' interests have some property $\phi$, but to make this the real answer to our question is to affirm that $\phi$-ness, not the satisfaction of Jones' interests, is the ultimate criterion. Suppose that Jones' interests were to change. Then what? If they were no longer $\phi$, should we hunt with Jones or run with the $\phi$'s? We would be unable to do both, and this first reply to our question bids us leave Jones for the $\phi$'s sake.

My discussion of this first answer has perhaps been overly abstract, and some examples should be mentioned in order to bring home the point, as well as to enable our forthcoming discussion of Reply No. 2 to carry its due weight. Suppose, then, that our prophet's reply of Type 1 about Jones is that we should satisfy Jones' interests because Jones is stronger than anyone else. This again is a "might makes right" theory, of sorts, and implies that if, perhaps per impossible, Jones should grow weak, we ought to stop taking his interests as critical for value. (The theologian might reply that Jones (i.e., God) cannot become weak, being infinitely strong. This is irrelevant. If strength is a defining property of Jones, then 'Jones' is no longer a proper name, and the theory is not of the type we are discussing.) Again, it might be said that we should regard Jones' interests as criterial because Jones can do Z better than we can. This makes Z-performing the ultimate criterion, and if Jones did not have this ability, his interests would cease to be criterial. And so forth. Any such answer will relegate the satisfaction of Jones' interests to a logical second place, behind whatever we pick for $\phi$. 

So let us turn to Reply No. 2, which is that we ought to regard Jones' interests as ultimately criterial just because they are Jones'. Now, this might be taken to be equivalent to saying that we ought to do so for no reason at all. But this would reduce the position to nonsense, in view of our analysis of evaluative language. Saying that we ought to do it, that it would be right or good to do so-and-so, is saying that there is a reason, and not that there isn't a reason, for doing it. Sometimes, indeed, the reason for doing something is "Just for the heck of it," which in a sense is doing it for "no reason at all." But this doesn't help. In the first place, "just for the heck of it" is automatically not
MORALITY AND UTILITY

a compelling reason. It would be absurd to say that you ought morally to do something “just for the heck of it,” for this “reason” implies that if you don’t want to, you don’t have to. Moreover, “just for the heck of it” does not mean “despite the fact that it will kill us,” or in general, despite the fact that it will dissatisfy our interests, but rather, that it might be fun, and at least will satisfy an interest in variety for its own sake. In neither case is this relevant to our present question, in which we are considering the proposition that the satisfaction of Jones’ interests is the ultimate criterion of right and wrong, bearing in mind that ‘right’ and ‘wrong’ are moral words, whose meanings have been partially analyzed in Chapters II and V.

Thus the “reason” offered amounts to, “because Jones is Jones, and nobody else.” But can the “fact” that you are you, that I am I, or that Jones is Jones, be the sort of thing which can logically be advanced as supplying a characteristic (reason) in virtue of which anything might be thought to be better or worse than something else? Let’s take an example. Suppose that we have two identical Ferraris before us, which we shall call ‘A’ and ‘B.’ (Note that we could just as well have called them ‘B’ and ‘A’ respectively.) Suppose that each will travel equally fast, wear out at the same rate, be equally comfortable, impress people equally, and so forth. There just isn’t any discernible difference between them. Under these circumstances, would it make any sense to say that A was a superior car to B? Of course not. Saying that A is superior to B is saying that there is something about A in virtue of which it is preferable to B, and ‘being identical with A’ does not indicate a difference between them, other than the fact that there are two of them. ‘Being identical with A’ can’t operate as a reason supporting a differential evaluation, because it isn’t a “characteristic” at all: we could equally have called A ‘B,’ in which case ‘being identical with B’ would have been the same “characteristic” as ‘being identical with A’ is now. This means that the “fact” in question simply amounts to the “fact” that it is identical with itself. And this is true of everything. So if this is what is meant, then if ‘being identical with Jones’ gives any reason for anyone’s doing anything, then ‘being identical with Smith’ or ‘being identical with Narveson’ would work equally well. Since everything is identical with itself, ‘be-
FOUNDATIONS

ing identical with itself logically cannot act as a reason for supporting differential evaluations. Insofar as it acted as a "reason" at all, it could only support identical evaluations. Yet a differential evaluation is what would be needed to support Jones' claim to have his interests preferred to everyone else's.

What follows from all this is, of course, that if my (or anyone's) interests are going to be advanced as ultimately supporting evaluations, then everyone's interests do the same. But this is all that we need, for it is precisely what utilitarianism asserts: that satisfaction of everyone's interests (that is, "everyone's interests equally," as shown in Chapter VII) must be the ultimate criterion, and hence the criterion of moral value (Chapter II). Given that interests are what support evaluations, the fact that they are mine rather than someone else's is logically irrelevant.

Is It a Verbal Trick?

This proof, which derives from Hare,24 in some respects and Gauthier25 in others (and indeed, goes back to Kant and to Mill as well), may have the appearance of sleight-of-hand. For example, critics may suppose that I have, after all, attempted to fish a value out of a fact. The argument, it may be thought, starts with the premise that everyone in fact regards his own satisfactions as good, and then attempts to derive the conclusions that the satisfaction of everyone is good. This is one of

24 Freedom and Reason, especially Part II.
25 Practical Reasoning, pp. 87-89, surely makes much the same point that I am advancing here, although it is not emphasized so heavily. That it is not, is due (I think) to differing attitudes between Professor Gauthier and myself. He apparently does not regard this as an argument supplying a "justification" of morality. Differences between us on this matter could probably be split fairly satisfactorily. I find the same principle stated with the author's usual pungent conciseness in Sparshott's Enquiry into Goodness, pp. 168, 169: "By arguing in terms of goodness one claims, however wrongly, that one is not simply seeking one's own way . . . the fact that I am I or you are you can never by itself constitute a ground of relevance." See also David Hume, Enquiry Concerning the Principles of Morals, Conclusion (New York: Liberal Arts Press, Bobbs-Merrill Co., Inc., 1957), p. 93.
MORALITY AND UTILITY

the errors imputed to Mill, who argued in something rather like that fashion (of Mill's proof, cf. below). But this is not the argument, although the assumption in question does play a part. The argument is that anyone who proposed to regard his interests, as such, as providing reasons in support of evaluations, is logically forced to admit the similar relevance of anyone else's interests. Thus the premise which any particular person uses, concerns his own interests, and asserts of them that they are to be regarded as criteria of evaluation, which is an evaluative, not a factual, premise. The statement that everyone does so regard his own interests, is relevant only because it reminds us that everybody is a party to the argument, and hence that it appeals to everyone.

What I have purported to prove is, in effect, that it is irrational to deny the principle of utility, if one regards one's own interests as supporting evaluations. The point of supposing that this argument is somehow a verbal trick or sleight-of-hand must be to call into question something about it. What? Not, I trust, its logic. Indeed, it would not be attacked as sleight-of-hand if the logic were unsound, for in that case, one could merely call it unsound and be done with it. What people really have in mind when they use this particular mode of criticism must be something else. And we may well suspect what it is: they doubt that this argument really will move people to be virtuous. This is doubtless true (though it might move some, and especially, might keep some from abandoning virtue). However, I have not claimed that the argument will do this. I have claimed only that it will move rational people to virtue, in so far as they are rational; which is just another way of saying that it is valid. People do not necessarily act rationally. But that an action is irrational is itself a criticism of it, and that actions contrary to the principle of utility are thus to be criticized is simply another statement of what I have set out to prove.

A more serious criticism, it might be thought, is that the truth of the conclusion depends upon that of the premise, and that people might wish not to accept the premise. This was, it will be recalled, that my interests (each person taking 'my' to refer to himself) are to be regarded as supporting evaluations, or in other words, that my interests ought (prima facie) to be satisfied.
Let us consider what denying this might consist in. It means, for one thing, that if you beat me up, I have no right to complain, in any sense of 'complain' worth noticing. For obviously, only justified complaints are worth noticing. If I scream or say 'Ouch! Quit it!' but deny that you ought to quit it because what you're doing is wrong, then nobody need listen. Only justified complaints deserve our attention.

The truth is that it makes no sense to refuse to apply evaluative expressions on behalf of one's own interests. This is merely a dodge. When a man says, "well, I for my part refuse to regard the satisfaction of my interests as good," he's being insincere, and the reason is very simple: he wouldn't say this unless he regarded his refusal as justified. He is trying to use his refusal as a way of impugning the validity of an argument, and this is to supply a counter-argument of one's own. But arguments about what to do cannot be carried on without the use of evaluative terminology. This is why the skeptic's position is incoherent at bottom, and why our hypothetical critic's position is in the same boat. The skeptic, when he says, "Nothing is really good or bad" (offering perhaps as his reason that evaluative terms are "meaningless" or "merely subjective") is attempting to offer us reasons for avoiding the use of evaluative terminology. If he is not, his argument is pointless, and he may just as well clam up. As Aristotle rightly says, the man who denies the law of contradiction is no better off than a vegetable. But to claim to be justified in refusing to use evaluative terminology is to apply evaluative terminology to one's refusal, and this is self-defeating.

In short, what the skeptic and our critic are really trying to say is this: one ought not to employ evaluative terminology. The absurdity of this remark needs only to be pointed out. There just isn't any way for anyone who is capable of talking about values to make out a coherent argument against the practice of employing evaluative language on behalf of his interests; and thus, there is no way for him to make out a coherent argu-

---

26 If 'argument' is taken in the epistemic sense, then perhaps this is true of all arguments, and not merely of arguments about what to do. Whether such terms as 'know' are at least partly evaluative is a vexed question. Cf., R. Chisholm, *Theory of Knowledge* (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1966), pp. 11-14, for example.
MORALITY AND UTILITY

ment against the principle of utility. This result is as much as it is possible for the theoretician to achieve; but it is also as much as he needs to achieve. For the reasonable person is precisely the person who acts on the basis of reasons, and those who do not can only be influenced by non-rational means, such as force. But they are also precisely the ones who cannot complain at its use, as we have seen.

Other Dodges: Hare’s Fanatics

In *Freedom and Reason*, R. M. Hare outlines a procedure for moral argument not unlike the foregoing, in point of the purely logical apparatus employed. Hare calls attention to the two essential features of evaluative discourse, that it commits people to action, and that it purports to supply reasons, which (therefore) must be of a general kind. But Hare’s account differs importantly in one respect. Instead of employing merely the general principle (if that’s the right word for it) that my interests support evaluations, he proposes that we apply a universalization method to each particular one of our proposals for action. His procedure is thus reminiscent of Kant’s, in which we are supposed to universalize the maxim of our acts. If its universalization is consistent, then action on it is morally permissible; otherwise not. 27 The defects of this procedure are the same for

27 I have not emphasized the differences between Kant, Hare, and myself on this matter of “universalizability” in the text, but it should be pointed out that none of us is to be identified with any of the others. I have distinguished Singer’s procedure, singled out for criticism in Ch. V, which consists in asking whether it would be disastrous or undesirable if everyone were to perform the sort of thing in question. Hare, on the other hand, usually asks the question, How would I like it if I were in the other man’s shoes? He is, in other words, mainly concerned with “reversibility,” as Baier calls it (*Moral Point of View*, pp. 202-03). This has been criticized interestingly by Alf Ross in *Danish Yearbook of Philosophy* (Copenhagen: Munksgaard, 1964), in his article, “On Moral Reasoning.” Kant’s view, according to which a necessary (and sufficient, I think) criterion for the permissibility of a principle of action is that it be “capable of being made a universal law” has been interpreted along Singer’s and Hare’s lines, by various people, but doesn’t seem to me to be quite identical to either, though more similar to Hare’s than Singer’s, certainly. My view seems to
both Kant and Hare, and are worth going into here; they apply also, with some reservations, to Singer's method. First, if my interest is in, say cultivating roses to the exclusion of all else, we shall get results which all of these authors would concede were irrelevant; namely, that for extraneous reasons having to do with the nature of things and the constitution of the human frame, we cannot all cultivate roses to the exclusion of all else. Singer, as we have noted, patches this up by talking of "invertibility" and "reiterability"; Kant, I believe, does not even recognize the difficulty, and his interpreters have had to patch him up as best they can. 28 Hare's moves are somewhat similar to Singer's. All of this complicated logical maneuvering is unnecessary, of course, on the utilitarian theory. If I plant roses, there is no reason to think it will harm anybody, and so there's no problem. It's as simple as that. Second, and related to it: some people might like doing things which other people dislike. Now Hare's formula goes:

If, when we consider some proposed action, we find that, when universalized, it yields prescriptions which we cannot accept, we reject this action as a solution to our moral problem—if we cannot universalize the prescription, it cannot become an 'ought.' 29

Of course, the consequences which I can accept might be quite different from what others can accept. Thus, if I happened to like being in prison, the fact that upon universalization of a certain line of action ("let all debtors be put in prison") I would find myself in prison (being a debtor), might not bother me at all. Hare recognizes this, and to a certain extent sees the consequences. In his chapter on "Utilitarianism," he takes up Braith-

28 See Paton's article referred to in Ch. IV, n. 11.
29 Freedom and Reason, p. 90.
waite's example of the chap who likes classical music and lives in a thin-walled apartment, next door to a man who plays jazz on his trumpet.

Now it is obviously of no use for B to ask himself whether he is prepared to prescribe universally that people should play trumpets when they live next door to other people who are listening to classical records. For if B himself were listening to classical records (which bore him beyond endurance) he would be only too pleased if somebody next door started up on the trumpet.\(^{30}\)

The solution is, of course, that

B has got, not to imagine himself in A's situation with his own (B's) likes and dislikes, but to imagine himself in A's situation with A's likes and dislikes. . . . The natural way for the argument then to run is for B to admit that he is not prepared to prescribe universally that people's likes and dislikes should be disregarded by other people, because this would entail prescribing that other people should disregard his own likes and dislikes . . . It does not follow from this that he will conclude that he ought never to play the trumpet when A is at home, only that he will not think that he ought to have no regard at all for A's interests.\(^{31}\)

Now this comes very close to my argument above, except that the conclusion is properly to be stated that he ought to have equal regard for A's interests, it remaining true that this does not make the precise solution in any given case self-evident. What I want to point out is that this general result is really implied from the beginning, and renders unnecessary the employment of the "universalization" procedure on each particular problem. The same, precisely, applies to Kant. If what sets the whole stage for moral questions is the universalizability of my inclinations (desires, "willings," wants), then the result we inevitably are going to arrive at, if we commit no logical blunders and introduce no irrelevant metaphysical considerations about the "purposes of nature" (as Kant does), is utilitarianism, i.e., the position that the essence of morality consists in regarding any desire, no matter whose, as \textit{prima facie} worthy

\(^{30}\) \textit{Ibid.}, pp. 112-113.

\(^{31}\) \textit{Ibid.}
of satisfaction. We can then simply apply the principle of utility to particular cases. This is a much better method because it does not get stuck in particular cases by concentrating on just the one desire or set of desires that happen to get it going, but also all incidental desires bearing on it at the time; and it immediately bids us see that the fact that one person's desires are different in structure from another's is only to be expected, and not a special source of theoretical difficulty calling for hedging devices.\(^{32}\)

Why isn't Hare satisfied to rest with his method as a proof of utilitarianism, then? Because of the possibility of people he calls "fanatics." The fanatic is the person who is willing to take any consequences of the universalization of his desire. Thus if he is a fanatical anti-Semite, who holds that all Jews ought to be put in gas chambers, he might, upon seeing that this implies that if he were a Jew he too ought to be gassed, accept this consequence. "Logic," as Hare says, "does not prevent me wanting to be put in a gas chamber if a Jew."\(^{33}\) The fact that this would be contrary to my interest, were I a Jew, he regards as indecisive, for:

A person who was moved by considerations of self-interest and was prepared to universalize the judgments based on it, but had no ideals of this fanatical kind, could not think this; and it might plausibly be said that a man who professes to think this is usually either insincere or lacking in imagination—for on the whole such fanaticism is rare. But it exists . . . His [the fanatic's] ideals have, on the face of it, nothing to do with self-interest or with a morality which can be generated by universalizing self-interest; they seem much more akin to the aesthetic evaluations discussed in the last chapter. The enormity of Nazism is that it extends an aesthetic style of evaluation into a field where the bulk of mankind think that such evaluations should be subordinated to the interests of other people.\(^{34}\)

\(^{32}\) The devices in Singer, discussed in Ch. 5, may be so classified. It is well-known that Kant ties himself into knots on the failure to make any allowances for conflicts of principles conforming to the Categorical Imperative (see his "On a Supposed Right to Tell Lies from Benevolent Motives," in the Appendix to Abbot's edition of Kant's Ethical Writings [New York: Longmans, Green, and Co., 1873, 1954]). He would get into similar difficulties here.

\(^{33}\) Hare, Freedom and Reason, p. 110.

\(^{34}\) Ibid., p. 161.
This is a powerful-looking objection, for if Hare is right, then contra-utilitarian theories could be logically coherent after all. But I believe that Hare is mistaken about the role of the expression ‘self-interest.’ In believing that it is possible for a person not to be “moved by self-interest,” he mistakes the “principle” as I called it above, for a genuine substantive hypothesis. My contention urged in Chapter III and elsewhere, is that this supposition that people are self-interested is wholly vacuous; and part and parcel of this is the corresponding view that the “assumption” discussed in the section above, that people employ evaluative expressions in support of their interests, likewise is vacuous. It is not just “plausible” that a man who professes what Hare’s Nazi professes (professes to profess?) is insincere: it is certain. Or in other words, the Nazi is confused, if he is “sincere.”

Why? Let us recall that I too have contrasted aesthetic with moral evaluations (Chapter II) and have admitted, or even insisted, that there are questions of “intrinsic value” of which aesthetic questions are either the best examples or, perhaps, the only basic examples; and that these questions are perfectly genuine. But none of this matters. It is either true or false, perhaps, that the world would be intrinsically best if it were $\phi$; and let ‘$\phi$’ be such that the gassing of Jews is an integral part of any world characterized by it. But my reason for acting on this cannot, logically cannot, be that such a world would be intrinsically best. My reason for acting on this cannot, or rather, what makes it my reason, is that I think it would be best. What I think when I think it would be best is, of course, that it would be best; but it is logically impossible for it to be the case that its simply being best is what (rationally) moves me to act. Only the belief that it is best can do this. Of course, I might get so involved with this belief that I might allow the interest in making the world conform to it to dominate all of my other interests, including those which keep me alive. That is what fanaticism is: it is having some interest so intense that no other interests count; one is willing to die for it and will not hear contrary reasons. But this does not necessarily mean that all fanatics are irrational. Doubtless they are, in fact, but we can conceive of a rational fanatic, or perhaps, someone just like a “fanatic” except rational. Such a person would be willing to die for the realization of his domi-
nant interest and would do absolutely anything to see it realized, except something immoral. For he would realize that his reason for acting to satisfy his interest is that it is in his interest, and if he wants to defend his activity on that ground, then he has to give equal weight to other people’s interests as well. It is true that people generally abandon their ability to reason morally as well as prudentially when they become fanatics. This might, as I say, be part of what is meant in calling them ‘fanatics.’ But this verbal issue is unimportant here, for Hare is not trying to defend irrationality, but to claim that fanatics might be consistent in their fanaticism because they might be willing to abandon self-interest. If my argument is correct, the expression ‘abandoning self-interest’ is logically inapplicable, except in the secondary sense that one might throw prudence to the four winds, i.e., might subordinate all of one’s other interests to one or a few which then dissatisfy the others (e.g., lead to death or permanent disability). But here Hare’s stricture is appropriate: we call such people imprudent, but a person might regard one of his interests as superior to all others, and there’s nothing for us to do about it except try to talk him out of it (or advise him to see a psychiatrist).

I conclude that fanaticism is a dodge, just as the effort to refuse the use of evaluative expressions on one’s own behalf is a dodge. Every effort to escape morality is a dodge, in fact, just as, and for the same reason as the effort to escape “rationality” is. We hear of people who profess to be against rationality, in some sense or other. The futility of this is transparent: all of these people employ words like ‘therefore,’ ‘not,’ and the rest of the vocabulary of reason, and could not state their “views” without them. But all that this shows is that the only way to do without rationality is to join the vegetables, which must perforce remain silent.

*Mill’s Proof*

The fourth chapter of Mill's *Utilitarianism* has become a classic example, though usually regarded as a classic bad example rather than a classic good one, for beginning students of
philosophy. In a way, I suppose, this is unfortunate, for it is not very easy to assess the effort as a whole, and not at all obvious that the famous blunders of which he is accused have actually been committed. It is perhaps the duty of an advocate of utilitarianism to add his bit to the understanding of the argument Mill supplies, and I shall try to do so here very briefly and without much detailed attention to the opinions of other critics.\textsuperscript{35}

As we all know, there are thought to be two crucial turns in the argument. The first of these is the famous parallel between ‘desirable’ and ‘audible’:

The only proof capable of being given that an object is visible, is that people actually see it. The only proof that a sound is audible, is that people hear it: and so of the other sources of our experience. In like manner, I apprehend, the sole evidence it is possible to produce that anything is desirable, is that people do actually desire it.\textsuperscript{36}

Mill goes on to assert that people do actually desire happiness, and indeed nothing but happiness, the critics nipping at his heels all the way through. What are we to say?

The odd thing here is that what Mill actually asserts is unquestionably true: the only proof you can give that something is desirable is that it is desired (or anyway, would satisfy a desire). Any serious amount of attention paid to the way in which arguments about the desirable go will verify this point, and unsurprisingly so.\textsuperscript{37} The only question is whether what we are trying to prove when we try to prove that something is desirable is the same sort of thing as what we are trying to prove when we try to prove that something is audible, allowing for the differences between desiring something and hearing it. (At


\textsuperscript{36} Mill, \textit{Utilitarianism}, p. 32.

\textsuperscript{37} This is well put in Arthur Murphy’s book, \textit{The Theory of Practical Reason}, p. 44, wherein he says, “The question is . . . what is it reasonable for me to do?” It is this latter question that cannot be sensibly asked or answered without a reference to the “thing wanted . . . without an intrinsic reference to our wants there could be no practical reasons.”
least one important author, Sparshott, claims that it is, and his argument is worth serious consideration.) 38 Moreover, the common contention that ‘desirable’ means ‘worthy of being desired’ rather than ‘capable of being desired’ does not cut as much ice as needs cutting here. 39 For one thing, ‘capable’ is a little thin in the nonevaluative contexts. ‘Audible’ surely means ‘apt to be heard (under normal conditions)’ and not just ‘capable’ of being heard (by whom? Gustave Mahler, or the rest of us?). And desirable things are apt to be desired by those who take the trouble to acquaint or interest themselves in the kind of thing in question, and could not intelligibly be held to be desirable if they weren’t. We do not, as Sparshott justly points out, pay any attention to the desires of parties who just aren’t interested, when it comes to talk of ‘good’ (‘desirable’). 40

The other major move is from “each person’s happiness is a good to that person” to “the general happiness is a good to the aggregate of persons.” People have accused Mill of a fallacy of composition in this passage, but this is an implausible accusation. In fact, the main thing puzzling about this move is in the ‘to’ phrases. In “each person’s happiness is a good to that person,” ‘to’ is clear enough: it just means that he thinks it a good. This I have endeavored to support on logical grounds above. But in “the general happiness is a good to the aggregate of persons,” ‘to’ is puzzling, since the aggregate is not capable of having an opinion on this unless it is meant that each person thinks the general happiness is desirable, and of course that may or may not be true but doesn’t follow from the premise, and is not part of his argument anyway. Nor is it what Mill needs, which is why the “fallacy of composition” interpretation is unsatisfactory. For surely what Mill meant to try to prove is that the general happiness is a good, period. Moreover, he says so in one of his letters, where he declares that he meant merely that if A’s happiness is good, B’s good, and C’s good, then the happiness of all of them (or the “sum” of them, i.e., the “general” happi-

38 Inquiry into Goodness, Ch. 5 and Ch. 6, especially.
39 The common contention in question is advanced in Principia Ethica, No. 40, and (in ref. to Dewey rather than Mill), by Morton White in his “Value and Obligation in Dewey and Lewis,” in Sellars and Hospers, pp. 332 ff.; and in other places too numerous to mention.
40 Inquiry, pp. 143 ff.
MORALITY AND UTILITY

ness) must be a good.\(^{41}\) Only this confuses the issue because in
this explanation, he has the premise different from what it is
in the passage from *Utilitarianism*, which has it not that A's
happiness is a good, but that it is good to A, i.e., A *thinks* it good.
Thus we must admit that Mill was apparently muddled about
this. But the argument isn't. As we have seen, the premise from
which the proof of utilitarianism proceeds is precisely that each
person thinks his happiness a good, and the conclusion is pre-
cisely that the general happiness is a good, and the conclusion
does follow, in the required sense and as a matter of strict logic
rather than as a matter of psychology. As to its being the only
good, on the ground that happiness is the only good, we have
seen what to say about all of this.

There are a number of peculiarities, certainly, elsewhere in
the chapter. The most peculiar thing of all is the statement that
people desire happiness, understood as being parallel to the
statement that they desire money or virtue. Some of the pecu-
liarity is dispelled by Mill's qualifications and explanations,
though not all of it. Critics have trodden very heavily on Mill's
discussions about virtue becoming "part of the end," though not
originally so, and here again I think that Mill may be interpreted
as making perfectly sound points, though in a rather confusing
way. He says, e.g., that

Virtue, according to the utilitarian doctrine, is not naturally and
originally part of the end, but it is capable of becoming so; and in
those who love it disinterestedly it has become so, and is desired
and cherished, not as a means to happiness, but as a part of their
happiness.\(^{42}\)

All of this is acceptable, if we sort things out a bit. To begin
with, when he says that virtue is not originally part of "the"
end, we have to take him to mean that people do not originally
in fact aim at it, not that it isn't something they *should* aim at,
since obviously every moralist wants to defend the latter. Un-
derstanding him in this way, what he says is quite true: you in


\(^{42}\) *Utilitarianism*, p. 34.
general have to *train* people to be virtuous, and there are probably no exceptions at all to this, though there is much variation in the degree and kind of training needed.48 Secondly, there are two kinds of good people: those who enjoy making other people happy (and thus include the activity of making them so as "part of their happiness"—part of the life they see as a good kind of life to live), and those who perform their duties grudgingly, in order to ward off attacks of conscience, or with a view to public acclaim and reward. Both of these latter types regard virtue as a means to their own happiness, not in the sense that they enjoy it, but in the sense that they get something out of it which they do enjoy (or at least relief from the pangs of conscience for bad performance). The analogy with money is to a degree apt here, but I think Mill doesn't quite appreciate its aptness, for in fact to say that virtue is desirable is just as misleading as to say that money is desirable. Both of these are inherently worthless, in one sense: virtue just consists in doing good, and it is what is done that is good, morally speaking, not the activities apart from their consequences; similarly for money, as Mill says. But on the other hand, to say that they are "inherently worthless" is a bit misleading too, for virtue is inherently a matter of trying to achieve certain consequences, and money is inherently a means of exchange: in neither case are the things just accidentally related to their effects, but rather, are so named because of their intended effects. The difference is that it is silly to get a fixation on money apart from what it will buy, whereas there is something to be said for people who pursue virtue "apart from its consequences" (since it cannot be pursued apart from its consequences, whereas money can: you can accumulate heaps of buying power without spending it, but you can't do heaps of virtuous acts without making people happy, or at least by preventing unhappiness).

But Mill does not clearly realize that the principle that people "desire happiness" is merely verbal, and this is why his efforts to prove this are strained and sometimes inappropriate. I attribute this failure to the general intellectual atmosphere of the time. Only recently have we begun to see through some of the psy-

48 This point, familiar since Aristotle, is well reinforced by Nowell-Smith, pp. 255–59; in a somewhat different way by Baier, Ch. 7.
chologisms of the last century, as in reality “logicisms.” The proof is not very much the worse for wear for that, for the idea behind it all is sound.

The truth is, that the best way in which the proof could have been stated is scarcely bothered with. I refer to the passage in which Mill says, “Happiness has made out its title as one of the ends of conduct, and consequently, one of the criteria of morality.”44 Very little attention is given to this ‘consequently.’ In the first chapter, he does say that “All action is for the sake of some end, and rules of action, it seems natural to suppose, must take their whole character and colour from the end to which they are subservient.”45 In Chapter II, he says in reference to an immediately preceding discussion of happiness, “This being, according to the utilitarian opinion, the end of human action, is necessarily also the standard of morality.”46 The idea being expressed in these few meager places is certainly the central idea of the whole utilitarian viewpoint: that the purpose of morality is to forward the ends which people severally have, whatever they may be. If Mill had concentrated on this way of putting it, instead of the rather obscure and “psychologistite” way he does, he probably would not have been the subject of nearly as much fuss as he has been. But what we have is a great deal better than nothing, and indeed not really wrong at all, though inadequately explained and clarified. Thus, at any rate, he may be interpreted.

On the Nature of Practical Reasoning

My arguments have purported to show that it is irrational to deny the principle of utility; and I have also claimed that it makes no sense to suppose that there is an open question whether one ought to be reasonable. The net effect of these arguments, then, would seem to be that rationality is somehow logically linked to sociality. The dice of reason are loaded in favor of the

general interest. This seems to require an explanation of some sort, one which would clarify the connection between reasoning in practical matters and society’s interests. Let us consider how such an explanation might proceed.47

Certainly the most popular philosophical theory about the nature of reason, until recent times, has been the one deriving from Plato, and occurring, in more detailed form, in Kant. According to this general type of explanation, Reason is a Faculty of the soul, one among others. In the case of Plato, there are two others, Appetite and Will, while in Kant only one is relevant, Inclination. The soul is pictured as being capable of internal conflict, Inclination issuing one set of orders and Reason issuing another, often contrary, set. Views of this type have a certain charm, no doubt, but they simply won’t do. The overwhelming objection to them is that they don’t explain what they set out to explain. If the commands (or advice?) of Reason are so characterized simply because they happen to come from Faculty X rather than Faculty Y, then the question why we ought to obey one rather than the other becomes open. There is then no answer possible to the question, Why does the Faculty of Reason issue the particular commands it does, rather than some other commands? For in saying that ‘Reason’ is simply the name of this faculty, all we can say is that Reason just does issue those commands and no others. Nor will it do to go on explaining, in a sort of proto-biological fashion, “why” Reason issues those commands (“Well, you see, Reason happens to be equipped with the Principle of Non-Contradiction, the Law of Causality . . .”: compare with, “Well, there’s this cog and it turns this wheel, which moves that lever, and thus . . .”). What we want to know is whether those commands are reasonable, regardless of who issued them. The explanation of rationality in terms of an appeal to a Faculty of Rationality, then, accomplishes noth-

47 Baier discusses to good effect some suggestions about this, pp. 89-92, and is very good on the motive power of reason, pp. 147-48. But I do not think he quite answers to our satisfaction the question raised here of the relation between the “essence” of reasoning and the resulting bias toward the general interest. I think this is in part because that “bias” is not (as we saw in a previous section of this chapter) so thoroughgoing as I make it out; and partly because, in the end, there is no real difference between Baier’s own view and the view of Hume as he presents it.
Morality and Utility

ing. We should just have to engage in discussion with the Faculty of Rationality to see whether its commands were justified, and this is what we wanted to know in the first place.\footnote{There is a clear connection between this view of Reason, and views which put the ultimate authority of ethics in God, or in Conscience, or indeed anywhere; for in form, this is an authoritarian view. "Reason" is treated as a sort of abstract and ghostly authority.}

For precisely the same reason, it will not do to fix on some plausible principle and proclaim that this "just is" one of the defining principles of Reason. This move also seems to make the whole thing arbitrary. If "reasonable" principles just happen to be this particular set of principles rather than that, then the question why we ought to follow them again seems to be open.

This is difficult territory, and I offer the following account in a spirit of exploration; perhaps, in the end, it will prove no more satisfactory than the others.

Reasoning, to begin with, is a verbal activity; it is also, by and large, a social one, and unsurprisingly so, since we need to be trained to talk by those who already know. Why engage in it? In particular, what are we doing (i.e., what are our purposes) when we reason about practice?

To reason is to criticize and rebut criticisms. When we raise the question, "Why?" in any context, whether theoretical or practical, we envisage the possibility that something may have gone wrong. The supposition is that whoever supports the stand, the assumption, or whatever it is which is being called into question, is attempting to do a certain thing and possibly failing. In the case of actions, to ask "Why?" is to call for a justification of the action in question. The possibility that rectification is needed is implicit in the asking: those who are supremely confident do not ask it, except perhaps in a purely rhetorical way.

But also, questions have an audience, sometimes (usually) social, but sometimes consisting only of ourselves. Still, it doesn't matter how large the audience is, for to level a question at a particular audience is automatically to make them parties to the questioning. It would be pointless for me to raise a question before an audience if I had no intention of listening to their reactions or taking account of them if they had them; and pointless
to raise it at all, then, if no possible audience's reactions are envisaged as being worth taking into account.\textsuperscript{49}

To engage in the defense of a line of action, now, is to imply that the members of one's audience have certain characteristics which make their opinions worth taking into account. It is, in particular, to assume that they can understand one's presentation, and are able to respond with intelligible criticism. But this in turn means that one concedes a potential (at very least) of practical reasoning on their parts. To see, in the light of this, why “the dice of reason are loaded in favor of the general interest,” we can reflect that it would be absurd (i.e., pointless) to raise questions and make defenses of one's acts to other people, if one were not prepared to acknowledge similar weight to similar claims on their part. If I presume to defend what I am doing before an audience consisting of Jones by saying, “Well, I enjoy this sort of thing,” but am not prepared to allow Jones to defend what he is doing with the same sort of argument, then I might as well not have bothered. It makes no sense to address people in the terms of reason if one isn't prepared to grant them any status as reasonable beings. This, I conceive, is the dialectical problem facing proponents of nondemocratic forms of government: arguments on their behalf can hardly be addressed to everyone in general, with a straight face. The same is true of moral arguments. There is no point in being prepared to argue if one doesn't envisage any possible terms of settlement; and one cannot hope for settlement if one's “arguments” are going to be arbitrarily loaded in favor of oneself.

Those who would retreat into their shells and argue only with their alter egos, however, gain no advantage from the tactic. They either must use, as their reason for refusing to face their fellow men in argument, the assumption that all other men are fools—an assumption which has to be put to the test before it can be accepted with any pretense of rationality—or must eschew the use of “reasons” altogether, in which case any monologues they may care to indulge in likewise abandon any pretense

\textsuperscript{49} The paper by Haworth, mentioned in Ch. VII, n. 19, makes use of similar considerations. See also some of the thoughts of Bernard Mayo in his book, \textit{Ethics and the Moral Life} (London: Macmillan Co., 1955), especially Ch. IV.
to rationality. Even in talking with oneself, one assumes that one is capable of understanding one's own arguments, assessing them, etc.: how, then, can one defend a refusal to listen to others who have the same capabilities, or worse, a refusal to find out whether they have them?

The answer, then, to our question why rationality in practical matters leads inexorably to acknowledgement of the general interest as the standard of judgment is that to raise a critical question about behavior, and especially to defend one's own behavior, is to start an argument; to start an argument is pointless if it cannot be settled; and it cannot be settled unless all parties to it have the same arguments available to them.

This leads us to formulate the argument for utilitarianism in an alternative way to the rather abstract-sounding one pronounced in the foregoing pages. To argue for morality at all is to claim the assent of all rational beings. But the only principle mutually acceptable to all rational beings is one which regards all of their interests as equally worth satisfying (less, therefore, those which are incompatible with others). Everyone can agree to this because everyone's interests are respected. No other view of morality can claim universal acceptance unless by accident. If the alternative morality involves recognition of some particular type of activity as intrinsically superior, or of some particular person as intrinsically superior, then it automatically debars itself from the assent of anyone who doesn't agree with that evaluation. And on the other hand, the utilitarian principle implies the desirability of acts based on the recognition of those superiorities if everyone recognizes them, since in that case, it will ipso facto be in everyone's interest to perform those acts. Thus in the event that conditions for universal acceptance of some logically alternative position are met, the principle of utility will automatically sanction acts based on that position anyway.

In the light of the above, it becomes still clearer why the depiction of rationality as one agency of the soul among others, whose business is to deliver a particular set of (unpleasant) orders, is unsound. Anyone capable of handing down orders is, especially if he is worth listening to, going to have to be capable of giving reasons for them: thus anybody, or any agency for
that matter, purporting to be Reason incarnate, automatically invites derision. If he (or it) purports to have reasons for his orders, then he (or it) has to concede the equal claims of the other parties to the discussion, while if he (or it) has no reasons and claims that his orders are to be obeyed without question, his pretension to being the incarnation of reason is exposed as a sham. Conscience stands unmasked as just another one of the boys. Perhaps, in the end, all that Kant meant with all of his talk of Faculties and Transcendental authorities is what we have said above: that there could be no point in reasoning unless one was prepared to concede equal weight to the interests of all the participants in the discussion, and consequently, if we conceive of everyone as being in on it, of everyone. In the end, then, there is no genuine difference between the position that morality is “founded on” pure reason, and the position that it is “founded on” a sentiment or feeling of sympathy, unity, or benevolence toward one’s fellow men.