In Chapter 4 of his essay *Utilitarianism*, "Of what sort of Proof the Principle of Utility is susceptible," J. S. Mill undertakes to prove, in some sense of that term, the principle of utility. It has very commonly been argued that in the course of this "proof" Mill commits two very obvious fallacies. The first is the naturalistic fallacy (the fallacy of holding that a value judgment follows deductively from a purely factual statement) which he is held to commit when he argues that "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... the sole evidence it is possible to produce that anything is desirable, is that people do actually desire it." Here Mill appears to hold that "X is desirable (=fit or worthy to be desired)"—a value judgment—follows deductively from "People desire x"—a factual statement. And the second is the fallacy of composition (i.e. an illicit transition from a statement about each several member of a collection to a statement about the collection as a whole) which seems to be involved in Mill's argument that since "each person's happiness is a good to that person... the general happiness (is), therefore, a good to the aggregate of all persons."^2

Attempts to defend Mill against these criticisms are considerably less common, but James Seth attempted to do so in 1908^3 and, more recently, Professor E. W. Hall^4 and Dr. D. D. Raphael^5 have done the same. I propose in the first part of this paper to consider the alternative interpretations offered by these writers, in particular those offered by Raphael as the most recent of them, of the passages quoted above; and in the second part to try and evaluate the alternative interpretation suggested by Seth and Hall of Mill's

---

1 *Mill's Utilitarianism*, etc., ed. Plamenatz (Blackwell, 1949), p. 198. All page references to Mill are to this edition. I shall refer to the paragraphs of Chapter 4, which contains Mill's supposed proof, by number—there are only 12 of them. The above quotation is from paragraph 3.

2 Ibid.


"PROOF" OF THE PRINCIPLE OF UTILITY

argument as a whole. I hope to be able to show that whilst the alternative interpretation of the whole argument is plausible and legitimate it is not, as his defenders claim or, at least, imply, much more favourable to Mill's reputation as a moral philosopher than the usual one.

I.

I think it beyond doubt that the passages quoted, taken by themselves, may very reasonably be supposed to contain the fallacies they are commonly alleged to contain. Mr. Weldon is surely unique in thinking it so obvious as to need no argument that Mill's admittedly "much-criticized" words in the first passage are "perfectly correct."1 But, as Mill's defenders point out, these passages do not stand alone and must be interpreted in the light of their context.

If this be done, it is argued, it becomes apparent that Mill is not, as the common criticisms assume, offering a strict (=deductive) proof of the principle of utility. That is to say, he is not arguing that if anything is desired then it is desirable and that happiness is desired therefore it is desirable. He repeatedly and emphatically states that a strict proof is impossible. Further, it is maintained, Mill does not think, as the criticisms also assume, that the passages against which they are directed contain a full statement of the sort of proof which he is offering. The paragraph in question is clearly preliminary. It is significant that it concludes with the very modest point that "happiness has made out its title as one of the ends of conduct."

All that Mill is there trying to do is to present certain preliminary considerations which, though falling short of strict proof, are, he thinks, nevertheless capable, when taken in conjunction with other considerations yet to come, of determining the intellect to assent to the principle of utility. Accordingly the points made in the disputed passages are regarded as rather trivial ones. Raphael interprets them as follows: 1. "That the idea of the desirable arises from the experience of desire in the sort of way ('in like manner'), but not in the same way, that the idea of the visible arises from the experience of seeing," (op. cit., p. 348); i.e. "that we talk of the 'desirable' as a result of having desires," (p. 349). 2. That—here Raphael relies on a letter of 13th June, 1868—"since happiness is at any rate one of the objects of desire, it is desirable or good, and that this applies to the happiness of all since everyone desires his own happiness" (p. 350). Raphael admits that he is not completely satisfied that the wording of the second passage will bear this innocuous interpretation—I am not sure that the wording of the letter will either—though he appears to think that it is consistent with the general tenor of the argument of chapter 4.

WHILST, of course, agreeing that the passages quoted from paragraph 3 should be interpreted in the light of their context, I am unable to feel sure that Raphael’s interpretations are correct, or that it is compatible with any plausible interpretation of the trend of Mill’s argument to regard the points made as such trivial ones. For (1) an apparent inference from the desired to the desirable occurs again in paragraph 9 (p. 202)—a paragraph which cannot be dismissed as merely preliminary, but which must, on any interpretation, be recognized as central to such “proof” as Mill is presenting. It is true that paragraph 9 will, as I shall show below, bear a different and more favourable construction, of which a very charitable reader might recognize the foreshadowing in paragraph 3: but it is also true that in the very next paragraph—No. 10—Mill explicitly identifies desiring a thing with thinking it pleasant and thinking it pleasant with thinking it desirable—which is exactly the fallacy commonly attributed to him. And (2) the second of the passages quoted must, again on any interpretation, be regarded as an important step in Mill’s argument. For it is the only place, so far as I can see, where he makes the crucial step from the desirableness of the agent’s own happiness, which he repeatedly insists is not regarded by Utilitarians as the criterion of right conduct, to the desirableness of the general happiness, which is.

In spite of my doubts about the adequacy of the above interpretations of particular passages, I have a good deal of sympathy with the onslaught on the common interpretation of Mill’s argument as a whole. Any alternative interpretation which does not involve attributing to Mill the commission of obvious fallacies has for that reason alone much to commend it. In the next part of this paper I shall consider a suggested alternative which has at least as much support in what Mill actually writes as the usual one and which is, in my opinion, the only serious alternative to it.

II

This alternative interpretation is, as I have said, suggested by Seth and Hall in the articles referred to above, and up to a point worked out by them. Raphael does not himself offer an interpretation of the argument as a whole, although he makes certain observations about the character of it upon which I shall comment in due course.

If it is to be properly evaluated, the argument attributed to Mill by this interpretation must be carefully stated and considered in the light of what he says he is trying to prove.

In the introductory chapter of his essay Mill undertakes to prove, in such a manner as the case admits, the utilitarian formula. This is given on page 169 as:

160
"PROOF" OF THE PRINCIPLE OF UTILITY

(a) "Actions are right in proportion as they tend to promote happiness."

He explains that the happiness here referred to is the general happiness, not merely the agent's own. The proposition supposedly proved in chapter 4 is, however, a different one, namely:

(b) "Happiness is desirable, and the only thing desirable, as an end."

There are several indications that Mill regards these two propositions as equivalent—which they plainly are not. Nevertheless, it is true that a proof of (b) would amount to a proof of (a), provided it could be granted that:

(c) "Actions are right in proportion as they tend to promote things desirable as ends."

For (c), taken in conjunction with (b), entails (a). And (c) might be granted on the ground that it is analytic. This is indeed to beg a disputed question in favour of the teleologists, but for the present purpose it is convenient and, I think, legitimate to do so.

It should be noted that if (a) and (b) are related in the way described the happiness referred to in (b) must be, as it is in (a), the general happiness. And, also, that "desirable" as it occurs in (b) must be taken in a normative sense, as meaning "fit (or worthy) to be desired" and not, for instance, "capable of being desired." Despite the apparent comparison of it with "visible" and "audible" (i.e. "capable of being seen" and "capable of being heard") in paragraph 3 Mill clearly does not in chapter 4 understand the word in the latter sense. He is not there trying to prove that the general happiness can be desired. He thinks that he has already shown this in chapter 3.

The "proof" of the principle of utility appears to be given in paragraphs 9 and 10 of chapter 4. In paragraph 9 Mill claims that if human nature is so constituted as to desire nothing which is not either a means to happiness or a part of it, then "we can have no other proof, and we require no other, that these are the only things desirable." And in paragraph 10 he claims that the antecedent of this conditional is true. Paragraphs 1–3 contain observations about the sort of proof which is being offered, 4–8 make certain preliminary points which tend to support the claim that the antecedent is true, and paragraph 11, the penultimate paragraph of the chapter, is designed to answer a possible objection to it. The short last paragraph simply repeats that the principle of utility is proved if the antecedent is true. "Whether it be so or not" being "left to the consideration of the thoughtful reader."

The first question to ask of this "proof" is whether the conditional upon which it depends is valid. For brevity it may be formulated thus: "If happiness alone can be desired then happiness alone is
desirable." I have italicized the word "alone" in its two occurrences because its inclusion makes a crucial difference between this conditional and the obviously invalid conditional—"If happiness is (or can be) desired then happiness is desirable"—on which Mill's "proof" has been commonly supposed to depend. The conditional upon which the "proof" depends according to the present interpretation seems to be valid provided that something is desirable. The truth of this proviso is the presupposition of the whole argument of Utilitarianism, in which the question asked is not whether there be a criterion of right and wrong but what is that criterion.

The grounds for holding the conditional valid are these: That to say that anything is desirable implies that it is capable of being desired. Hence if anything can be shown to be not capable of being desired it follows that it is not desirable. Now the antecedent of Mill's conditional states that nothing is capable of being desired except happiness, whence it follows that if anything is desirable at all that thing must be happiness. Consequently, if the antecedent be true, Mill has proved, and proved strictly, that happiness alone is desirable.

It will be observed that the above interpretation has the merit of regarding the much criticized paragraph 3 as merely preliminary to the main "proof," and does not involve charging Mill with the commission of the naturalistic fallacy. On the other hand, though Hall somewhat surprisingly does not appear to notice it, it is open to the objection brought against the usual interpretation, namely, that it represents Mill as offering that strict proof which he has declared to be impossible. The force of this objection can, however, in my opinion, be exaggerated. It is true that Mill does seem explicitly to rule out any strict (=deductive) proof of the principle of utility. "Strict" is not in fact his own word—pace Hall who misquotes the first sentence of chapter 4 as asserting that a strict proof is impossible where questions of ultimate ends are concerned. But, of course, Mill does say that proof "in the ordinary and popular meaning of the term" or "direct proof" or "proof by reasoning" is impossible (p. 166 and p. 198). And it is at least natural to conclude from these remarks that the sort of proof which is appropriate to questions of ultimate ends, proof in the "larger meaning of the word," cannot be deductive, even though Mill also says that the considerations he thinks "capable of determining the intellect either to give or withhold its assent to the doctrine (of utility)" are "equivalent to proof." Nevertheless, when Mill actually gives his "proof," it does not, as Raphael suggests (op. cit., p. 346), consist simply in stating clearly the view to be proved and then meeting possible objections to it. Quite the contrary. Mill, as I have tried to show, presents his argument in deductive form, affirming a conditional and then claiming
that the antecedent is true. Raphael has, I think, very fairly described the procedure adopted to establish the antecedent. But granted the antecedent, Mill's claim seems to me clearly to be that the consequent follows deductively. Whether this can be squared with his frequent and insistent disavowal of the possibility of "what is commonly understood by proof" on questions of ultimate ends I am not sure. One should not be too ready to jump to the conclusion that Mill is inconsistent, but his defenders should not refuse even to contemplate this as an hypothesis. It is, however, possible—and the main passage (on p. 166) in which Mill disclaims the intention of offering a proof in the ordinary sense of the word will perhaps just bear this interpretation—that he is primarily concerned to insist, not that one cannot prove that happiness alone is desirable for its own sake by deductive argument, but simply that one cannot show this in the same sort of way as that in which one shows that e.g. listening to music is desirable—that is, by showing that it tends to promote that which is desirable for its own sake, namely, happiness. To say that happiness is the ultimate end is eo ipso to rule out the possibility of showing in this way that it is. But it is not necessarily to exclude the possibility of showing that it is by deductive argument.

The second and, in view of the affirmative answer to the first, crucial question to ask of Mill's "proof" is whether the antecedent—"Happiness alone can be desired"—is true. Mill begins by claiming that people certainly do desire happiness—this is perhaps the main point that the unhappily phrased paragraph 3 is intended to make. No one would wish to dispute that they do desire their own happiness and, since it is desirable if possible to avoid attributing to Mill the commission of the fallacy of composition in the second part of paragraph 3, it must be assumed that it is their own and not the general happiness that he has in mind when he makes this point.

Let it be granted, then, that people desire their own happiness. The next step, as Mill explains in paragraph 4, is to show that they never desire anything else. He admits, of course, that they do desire things which, "in common language, are decidedly distinguished from happiness" (p. 199). Some, for instance, desire virtue. (By this he must be supposed to mean not merely that they desire that others should be virtuous, but that they desire that they themselves should be virtuous, i.e. desire to acquire a settled disposition to perform those actions which tend to promote the general happiness.) In order to meet this difficulty Mill argues that the desire for virtue is not, appearances notwithstanding, distinct from the desire for happiness. For if virtue is desired it must be desired either as a means or as an end, either as a means to or as an ingredient in happiness. And he goes on explicitly to state that the same may be
PHILOSOPHY

said of the desires for money, power, and fame. It begins to look as if there is no desire of which this could not be said, as if Mill is prepared to call anything whatever that is desired either a means to or a part of happiness—a suspicion which is confirmed by paragraph 10. There Mill at first claims that the question whether people desire anything other than pleasure is a "question of fact and experience, dependent, like all similar questions, upon evidence." He goes on: "It can only be determined by practised self-consciousness and self-observation, assisted by observation of others. I believe that these sources of experience, impartially consulted, will declare that desiring a thing and finding it pleasant, aversion to it and thinking of it as painful, are phenomena entirely inseparable, or rather two parts of the same phenomenon; in strictness of language, two different modes of naming the same psychological fact: . . ." and that to desire anything, except in proportion as the idea of it is pleasant, is a physical and metaphysical impossibility" (p. 202, my italics).

It should not escape notice that the proposition "happiness alone is desired" undergoes a change of logical status in the course of the above argument. To begin with, by appealing to experience and observation, Mill appears to regard it as synthetic. The appeal to observation is appropriate only when it is a question of establishing the truth of a synthetic proposition, a proposition whose contradictory is logically possible. But at the conclusion of his argument Mill is to be found affirming that it is logically ("metaphysically") impossible to deny that happiness alone is desired. This is, in effect, to regard the proposition as analytic, as true in virtue of the definitions of its terms, quite irrespective of observation. So to change one's view about the logical status of a proposition is not, it is true, necessarily reprehensible. A proposition which at first sight seems to be synthetic may very well upon examination turn out to be analytic. Nevertheless it must be said that Mill's reference to observation and experience does tend to obscure the fact that his considered view is that the proposition in question is analytic. And this fact has, as will appear below, an important bearing on the cogency of his argument.

1 It is of minor importance that whereas he before spoke of happiness Mill now speaks of pleasure, for "by happiness is intended pleasure, and the absence of pain; by unhappiness, pain, and the privation of pleasure" (p. 169).

2 I have omitted from this quotation the clause "that to think of an object as desirable (except for the sake of its consequences), and to think of it as pleasant, are one and the same thing." Here Mill unquestionably expresses the view commonly attributed to him that "x is desired (=thought pleasant)" entails "x is desirable," as he also does on p. 226 note by saying that "happiness" and "desirable" are synonymous. It is very difficult to accommodate everything Mill says in a consistent interpretation.
Mill’s contention is, then, in the end, that it is analytic that everybody desires his own happiness. The conditional upon which Mill’s “proof” depends is valid. Its antecedent, understood as Mill understands it, is true because analytic. Does it then follow that the principle of utility has been proved, that the general happiness has been shown to be the one thing desirable as an end?

Clearly it does not. For, first, Mill has shown only that the agent’s own happiness and not the general happiness is the one thing desirable as an end. He has not shown, what he undertook to show and supposes himself to have shown, that an action’s tendency to promote the general happiness is the criterion of its rightness or wrongness. He has not even shown that a man who makes his own happiness his ultimate end, as he believes all men (logically) must, will in fact be led to perform those actions which tend to promote the general happiness. He may well have held, as many moral philosophers have, that the course of life which best promotes personal happiness is also the course of life which best promotes the general happiness. His discussion of the desire for virtue (cf. p. 201 especially) suggests that he thought that a prudent man would desire to be virtuous—i.e. aim at the general happiness—in the first instance as a means to personal happiness, although in due course he might come to desire it for itself—i.e. as a part of happiness. This view, of course, assumes that being virtuous is a means to personal happiness—a proposition which is perhaps broadly true in a well-ordered society, but which is certainly neither obviously nor universally true. If Mill’s argument depends on the truth of this distinctly disputable proposition it clearly falls far short of being in any sense a proof of the principle of utility. And, moreover, such an argument would hardly yield the conclusion that conduciveness to the general happiness is the criterion of right and wrong. At best it would show that this is a property common to all right actions—the criterion proper being conduciveness to personal happiness. And this Mill repeatedly insists is not the utilitarian view.

James Seth, in the paper referred to above, gives a similar interpretation of Mill’s argument, allowing or rather insisting that all it shows is that personal happiness is the only thing desirable as an end. He does not, however, appear to think that this is a ground for complaining that Mill’s performance falls seriously short of his promises. Without criticism he reports that Mill, having proved that personal pleasure is the only thing desirable or good, “seems to think that it follows that the good, and therefore the ethical criterion, is the general happiness, or the greatest happiness, not of the individual, but of the greatest number of individuals” (op. cit., p. 479). The ground for this conclusion, admittedly not made explicit by Mill, is alleged to be that “since pleasure is the Good, the
greater pleasure must, as such, be better than the less, and the
greatest happiness of the greatest number (if not of all sentient
beings) must therefore be better than the greatest happiness of the
individual or of any number of individuals less than the total number.
He assumes, as a matter of common sense, what Sidgwick represents
as the result of an application of the principle of impartiality or
equality, namely, that . . . each ought to count for one and no one for
more than one" (ibid.). Seth further contends, a little later on, that
the apparent contradiction between egoistic psychological hedonism
and altruistic ethical hedonism is resolved by Mill's doctrine of
sympathy.

This may or may not be a fair statement of what Mill had in mind
when he wrote the passages in question. But, in any event, Seth
does not sufficiently stress the fact that the implicit premises by
which Mill is supposed to effect the transition from egoism to altruism
are themselves moral principles. Hence, to employ them in the course
of a proof of the principle of utility is inconsistent with regarding
that as the ultimate moral criterion. The whole difficulty about
proving the principle of utility, as Mill recognizes but Seth forgets,
arises from this very fact that it is presented as the ultimate principle
of morality. Seth defends Mill against the common criticisms, but
only at the cost of attributing to him a line of argument which he
clearly repudiates. Perhaps Mill ought to have taken this line, as
Sidgwick did after him, and recognized certain principles as self-
evident intuitions. But that he did not is surely beyond dispute.

And, secondly, quite apart from the above deficiency Mill's sup-
posed proof is open to another fatal objection. Remember that,
given his terminology it is analytic that everybody desires only
happiness. It is not, of course, analytic in ordinary speech, as Mill
in effect admits when he allows that people desire things which are
in common language distinct from happiness. But, if it is analytic
that everybody desires only happiness, it follows that tending to
promote happiness cannot be the criterion of right and wrong, that
happiness cannot be the one thing desirable as an end (or, more
exactly, that there is no point in saying that it is the one thing
desirable as an end), for happiness is not in the relevant sense one
thing. Two men who both desire happiness do not necessarily desire
to bring about the same state of affairs, or even two compatible
states of affairs. They desire the same thing only in the sense in
which Francis I and the Emperor Charles V desired the same thing
when they both desired Milan—and were prepared to fight each
other for it.¹ It is an odd criterion of right and wrong from which

indebted to Professor A. E. Teale for drawing my attention to this interesting
and, I believe, conclusive passage.
it follows that any action which tends to bring about whatever any agent may desire is right, from which it follows that both parties to every quarrel may be right.

The above interpretation may very well be, I think on the whole that it is, closer to what Mill meant than the usual one. But I am not primarily concerned to recommend one interpretation above the other. What I am concerned to do is to combat the suggestion that Mill's supposed proof can be represented, not merely in a different, but in a much more favourable light than that in which it is usually represented. I have no wish to deny that Mill was a considerable moral philosopher whose writings merit serious and sympathetic study. But I am sure that in chapter 4 of *Utilitarianism* he was writing much below his best.

*University College of North Staffordshire.*