

UTILITARIANISM

The problem with which this paper¹ is an attempt to deal is wholly circumscribed and departmental and not by any means of the thorniest kind. It enquires into the meaning of certain kinds of propositions which frequently occur in the works of ethical philosophers, and so examines the method of investigation which they presuppose. It does not enquire, because it would not be particularly relevant to do so, whether the doctrines which embody these propositions are themselves true or false. It seems to me that expressions have been and go on being used by writers on ethics which look simple and harmless enough at first glance, but on examination break down hopelessly, and turn out to be not false but totally meaningless. This is serious because it naturally affects the validity of all that is thought to be deducible from or presupposed by them, and in fact vitiates much ethical writing beyond repair.

I have chosen to begin by examining the views of utilitarian philosophers because these seem to offer the clearest instances of what I mean. But my conclusions, if they are valid, apply beyond the limits of utilitarianism: precisely how far they reach will, I hope, become clear in the course of the ensuing discussion.

When in general we ask ourselves what is the meaning of such and such a word or sentence, the most natural method of discovering the answer seems to consist in asking the further question, 'How is the word used?', which in its turn would reveal what kinds of fact the word or sentence was being used to symbolise: in other words what, if it were the case, would be held to make the proposition or propositions asserted in the sentence true or false, probable or improbable. Words which are such that no situation conceivable by me would make the statements of which they are part either true or false, or more or less probable, are not significant, not

¹ Written c.1937? (All footnotes are editorial.)

UTILITARIANISM

properly speaking words at all, not elements of what I can call my language.

This account of meaning has sometimes been called the principle of verification; but I should prefer not to call it that, for reasons which I hope, presently, to make clear. Thus, when I ask what is the meaning of such words as 'green' or 'pillar box', it is clear that the sort of situation which is relevant to sentences embodying these words is a situation in which something is being, or could be or could have been, perceived with the senses. To say this is not, of course, either to assert, or to say anything which entails the assertion, that any one phenomenalist analysis of such propositions is correct: even if all forms of phenomenism are false, it is still true that perceptual situations, even if not they alone, are {strictly}² relevant to sentences embodying words like 'green' and 'pillar box', that it is certainly they, whether or not is they alone, that render the propositions which are the logical constructions out of such sentences true, or false, or probable.

Now when I come to ask a similar question about the meaning of so-called ethical words, or propositions containing ethical terms, the sort of situation which I find that I regard as relevant is one which consists of my awareness of something by introspection. Again, it does not *follow* either that ethical predicates cannot be used of anything save introspective data, or complexes at least one element of which is an introspective datum, or that some naturalistic analysis of ethics is necessarily the correct one; but it does follow that if ethical propositions purport to state anything or all – and this too has lately been seriously questioned – then introspective facts are at any rate among the facts to which ethical terms apply. Again, of all types of propositions those which on the whole have been least suspect, or, perhaps one ought rather to say, the propositions which have been regarded as clearest in meaning by common sense, are those which could be tested by direct appeal either to the senses or to introspection. This, in view of what we have said, is not very startling; consequently those theories of ethics which analyse ethical predicates as symbols denoting introspective data are *prima facie* easier to understand, even if they are not always more plausible than those which do not.

² { } = best omitted?

UTILITARIANISM

One of the ethical theories of this class which has long been regarded as at any rate one of the simplest and most intelligible views held by philosophers is the theory known as the utilitarian. According to at any rate most versions of it, it is the case that that and that alone is good as a means which contributes to the pleasure or the happiness of somebody; and that and that alone is good as an end which consists in or has as a characteristic of it {the} pleasure or happiness as a state of somebody's consciousness. I use 'pleasure' and 'happiness' deliberately as interchangeable terms since I take them both to denote characteristics of introspectible states of consciousness, or complexes of such states; and whether or not pleasure and happiness are different one from the other, it is the common character of the two words as denoting one or more than one verifiable characteristic of states of mind that makes them relevant to my argument.

If the notion of the good either is or entails the notion of happiness, then it may be said plausibly that the ideal, the greatest good, may be or may entail the greatest happiness: it may be asked whose greatest happiness? Bentham and James Mill, usually regarded as at any rate the most consistent holders of the theory, replied 'the greatest happiness of the greatest number of individuals'. What does this expression [*illegible*]³ mean? It is the main purpose of this paper to answer the question.

Here certain questions naturally arise which must be dismissed since their solution, however important, is irrelevant to this enquiry. The first of these is whether 'Good = pleasure' is a definition or a synthetic proposition of some kind: whichever it is, the expression 'the greatest happiness of the greatest number' must have some analysable meaning, whether or not the proposition that it and it only is the good is itself analytic. The second question, which, while it may seem so, is really no more relevant than the first, is this: is the good something which consists in, or at any rate is entailed by, only my own greatest happiness, such that the happiness of others, if it enters into consideration at all, enters it only as a possible means to promote my happiness, or is it⁴ a end in itself? The first is occasionally advanced by several forms of egoistic hedonism, the second is the differentia which

³ 'or p[]ing of it?'

⁴ sc. the happiness of others?

UTILITARIANISM

defines the so-called doctrine of universalistic hedonism. We are not concerned with the answer to this, for utilitarians at any rate, perfectly genuine question: for whatever be the answer to it, whether the greatest happiness formula describes the end or the means to the end, the formula itself must mean something, or at any rate parts of it must. Otherwise utilitarianism ceases to be true or false, and becomes unintelligible. Which seems *prima facie* unpalatable. Indeed it is sometimes taken for granted that Benthamite utilitarianism is, however false in fact, formally irrefutable. Our next task, then, is to ask the question: What if anything is meant by the expression 'the greatest happiness of the greatest number', and does it denote the kind of fact which can have goodness or any other characteristic significantly ascribed to it?

It is not my purpose, however, to attempt to inject life into a dead horse in order to flog it once more before you. That the expression in question is extremely confused if not actually *meaningless* as it stands has been proved many times over, and there is therefore no necessity to do so once more *in extenso*: in order, however, to proceed to my next point I must make the briefest possible statement about it. The proposition that the good consists in, or that it is my duty to promote, the greatest happiness of the greatest number entails that, given at least a certain finite quantity of goods, that is, means of happiness, to distribute between at least some finite number of persons, there is at least one possible way of distributing these which is such that as a result of it the greatest possible *aggregate* of happiness will *necessarily* coincide with the happiness of the greatest possible number of persons: or, to put it differently, that there must be some one method of action open to me such that it will necessarily be the case *both* that a greater aggregate happiness will result from it than from any conceivable alternative action, *and* that a greater number of persons will be made happier than they were before by it than by any alternative action. But the only reason for supposing that the two cases will coincide is that one somehow involves the other, because the sum total of additional happiness whose parts are the increments of happiness of each individual is necessarily greater the larger the number of individuals affected. Leaving aside what is meant by sum total, it is obvious that if the sum total = sum of increments, it may depend on the size of each increment as much as on the number of increments: so that by practising strict egalitarianism I

UTILITARIANISM

may bring about a sum total far lower than that which would be produced by aiming at inequality, that is by satisfying the desires of those who happen to possess the biggest appetites; their individual increments added together might very well (if such addition is conceivable at all) amount to a far higher total than would be produced by spreading the benefits I have to lavish more thinly, or, as some would say, more fairly. This amounts to saying that the coincidence of the case where there is the greatest sum total of happiness with the case where the greatest number of persons is made happier, if it even happened, would be purely fortuitous, and in any case cannot a priori be regarded as necessarily possible, and that therefore the very meaning of the greatest happiness of the greatest number principle as so interpreted turns out to rest on the false presupposition that the quantity of happiness is necessarily connected with the number of persons affected; that is, that there is some function which correlates the values of these two variables – that is, the variables vary concomitantly. Since this is seen to be not necessarily the case, the principle as so interpreted turns out to be in effect unintelligible.

Faced with this glaring difficulty, a utilitarian would have to modify his position and decide to describe it in terms of one variable only: either that of greatest happiness or that of greatest number of benefited individuals. The two versions would then run somewhat as follows:

(1) It might be said that the principle now is that that and only that is right or good or obligatory etc. which makes a greater number of individuals at all happier than any other, whatever the degrees of each individual increment. This, if happiness and unhappiness are suitably defined, seems logically self-consistent, and not demonstrably false, but is highly implausible, especially unless a certain assumption is made: for it would follow that in an imaginary case we were being recommended to do, or the good was defined as, whatever made a body of 1,000 persons even faintly happier, rather than turn our available resources to greatly increasing the happiness of another body of 900, who would otherwise, to make the case sufficiently paradoxical, continue in a state of acute agony. The assumption which makes this recommendation seem possibly a shade less wildly implausible is that it is likely to be the case that the increments of the 1,000, however small, would in most cases add up to a bigger total than the increments of the 900, however large: this is certainly what

UTILITARIANISM

Bentham seems to have had in mind when he formulated his position. I have just tried to show that this rests on an obvious mistake: and I do not think that saying it makes the position even seem more plausible. But even if it does do so, the point is that it brings us to the other alternative account, (2), which abandons consideration of the number of persons except in so far as it is relevant for measuring the sum total of happiness produced. This we must now consider.

(2) According to this view the greatest happiness of the greatest number principle asserts only that where mutually exclusive possible courses tend to produce different amounts of total happiness, that course should be adopted which is likely to produce the greatest amount. If there should be only one course open which was likely to secure that, then that course should be adopted, no matter how many or how few were made happier by it. But supposing there is more than one course likely to lead to the same highest total, then three choices are theoretically possible:

(a) of that course which leads to the highest average happiness, that is, when the amount of happiness divided by the number of beneficiaries produces the largest numerical sum, and if several courses do this, then it is a matter of indifference which is chosen.

Or (b) of any of the alternatives indifferently: all that matters is that the larger total is produced, never mind by which of the methods that do so equally.

Or (c) of that course which makes the largest number of persons at all happier than they were before, that is, where the denominator (?)⁵ is the largest compatible with each individual experiencing *some* increase of happiness.

(a) is highly inegalitarian, sacrifices the majority to the minority, and is rejected by Bentham; (c) is in fact what Bentham seems to have advised. We *should aim at* producing that distribution of the largest sum total which benefits more persons than any alternative course: and the expression ‘should aim at’ cannot, even if goodness can, be analysed into ‘do aim at in fact if we are guided by utilitarian considerations’, since that would be circular, but is strictly normative, that is, to be done because to make others happy, or to secure the greater total aggregate of happiness, is right, or good, or approved by us, or an end worth having for its own sake. This is a far more plausible position, and has been and is

⁵ IB’s query.

UTILITARIANISM

held by writers other than utilitarians, save that for pleasure or happiness they substitute other verifiable and quasi-verifiable states of mind such as virtue, or knowledge, or any other state or complex of states of human minds or situations involving these states as irreducible components. It is this apparently more reputable view that it is the purpose of this paper to undermine, since it seems to me that less obviously, but not less truly, it may be said to rest on a confusion which must, as soon as it is pointed out, be seen to be fatal to the whole position.

II

To restate the view which is being assailed, it holds that it is (*a*) possible (*b*) good, or right, or an end in itself, or the only ultimate end of moral action, or (some say) the only possible end of any human action, to act in a manner which is likely to bring about a larger total quantity of happiness, or of virtue, or of anything which is such that at least one of its constituents is a state of consciousness such that to it words like 'greater' and 'less' can significantly be applied; and, further, that where alternative courses are likely to result in the same sum total of this quality, if it be one, it is right to choose that alternative which results in its distribution among a number of individuals larger than would be affected by any of the other alternatives.

I do not for the moment ask whether it is good to promote such a state of affairs: let us see whether it is even possible; what the proposition which asserts the possibility implies. And by 'possible' I mean logically possible or thinkable: that is, is the state of affairs which the utilitarian programme purports to describe conceivable in any sense of the words?

These philosophers speak of the greater or lesser happiness of a number of individuals. Several questions seem to pose themselves as demanding an answer before we can proceed further: (*a*) What, if anything, do I mean when I say that someone else is now happy? (*b*) What do I mean when I speak of my own or someone else's happiness as being now greater, now less? (*c*) What do I mean when I speak of his happiness as a greater or less than my own? And finally (*d*): If I can say that X is happier than Y, can I in the same sense say that X is happier than Y + Z? And indeed can I say this at all, and if I can, what do I mean? To the answers to these questions the rest of this paper will be devoted.

UTILITARIANISM

(a) When I say that I was happy or experienced pleasure at a certain given date, I think that what I mean is either that I was acquainted by introspection with a characteristic of my then experience which I call pleasantness, or, if I was not, then in principle I could have been if I had taken the trouble to ask myself whether I was, that is, to introspect. I may mean more, but not less, than this: that is, if this is not part of what I mean, then I mean nothing; this is the only fundamental way in which I use the word 'pleasure'; the other uses are derived from it. If I attempt to convey the fact that I was happy to you, and you are unfamiliar with my language, and beg me to explain how I use the word, I cannot quite do what I should do if we were discussing some characteristic of material objects, that is, take you to see it or hear it and then point, since my introspective data are, as we say, private to me, and are not, as events at any rate, in any sense common to us both, parts of our public world. But as I assume, in supposing you to be a human being at all, that you are sufficiently like me to make it probable that you too can introspect, and sometimes do introspect, and that pleasantness may be a characteristic of your introspective data, I attempt to describe the material public situation which is such that I suppose that, if you are what I take you to be, you would, if you were in it, feel pleasure, and then beg you to reflect on what you would introspect, hoping by some such means to convey to you the meaning of my word 'pleasure'. But of course I can never be sure that you even begin to understand me, not because I have not made myself sufficiently clear, but because I cannot *know* that you do introspect, since I cannot verify the proposition 'You introspect' where 'you' is taken to stand only for that, though all of that, which I can verify by direct acquaintance, that is, the data – remembered, present to the senses, and anticipated – which constitute what I call your bodily appearance and certain correlated verifiable effects outside your body, which is all that I am acquainted with in the same sense as I am acquainted with my own body, or *my* introspective data. But although I do not know that you introspect, I can understand what is meant by the proposition 'Smith introspects', and even regard it as probable, without being able, even in principle, to verify it by acquaintance. The fact that I do understand this proposition seems to me indubitable: and yet I know that I do not mean by it either, as has recently been alleged, (a) sense data are occurring, or will occur, or could occur in my sense field at a certain distance from the sense

UTILITARIANISM

data caused by my body which will be followed by certain other sense data, all closely resembling the sense data which I should expect to see if I looked into a mirror, or expect to hear as proceeding from my own body when *I* am in a certain state, verifiable by introspection. For this would entail that the analysis of the proposition 'I am pleased', which includes references to introspective data, is different from the analysis of the proposition 'He is pleased', which cannot do so, so that the state of affairs described by the first is, as it were, categorically different from that described by the second. But if so, then I do not recognise the second description as a proper description of what I mean. Nor do I mean (*b*), which has been offered as a rival account: namely that I do indeed mean to affirm that the *same kind* of case is the case by the two propositions, but that, even when I say 'I am pleased', I can never significantly assert more than that I hear sounds, see sights, observe the pointers of measuring instruments, and so forth. So that the difference between 'I am pleased' and 'He is pleased' is the difference between two different sets of images, pointer readings and so on, that is, is the same difference in kind as the one between what is described by the phrase 'my state now' and the phrase 'my state a minute before'. Which comes to saying that neither you nor I can ever be *said* to introspect: for the word describes nothing not analysable into sense data.

If the first of these accounts is unacceptable because it denies the parallelism of the two cases by offering a behaviourist analysis of propositions about other selves, but a partially introspective analysis of propositions about my own states, the second is even more so: it realises that I do intend to assert the parallelism, that whatever the case of scientific propositions, the *common-sense* proposition 'Smith is pleased' *cannot* under any circumstances be translated or analysed into a set of behaviourist protocols without definitely altering the sense of the original proposition; it accepts the fact that common sense will not, if it knows what it means, be prepared to substitute the analysis for the analysandum, however many times it is assumed that analysis, being at a different level, cannot by its very nature alter the nature of the common-sense propositions which it analyses – and that because the question 'How can I be acquainted with *your* introspective data?' is asked and must be answered at a common-sense level or not at all. And therefore it resorts to the desperate remedy of reducing my proposition about myself to the level of those propositions about

UTILITARIANISM

others which I am sure I can verify: that is, of denying the existence of introspection-data, as some thoroughgoing behaviourists have done.

Against this one can resort only to the evidence of one's own experience, and refute the hypothesis by claiming direct acquaintance with the entities whose possibility it denies. This is my reason for earlier refusing to call the principle which I accepted for determining whether or not propositions were significant, the verification principle. I wish to hold that I can conceive of situations, that is, understand propositions asserting that there could be situations, which could not in principle be verified by me: it is in principle nonsensical to assert that X could verify Y's mind in the sense of being able to inspect Y's introspective data, that is, data numerically identical with Y's data and not merely highly similar data which might be obtained, say, by telepathy. Yet it is not nonsensical to say, and can be veridically asserted, that X's proposition about Y's introspection is in fact true: that is, in order to understand a proposition I do not even appear to need to have any cognitive attitude to the means of verification, let alone know them, or mean them by the proposition.

I am not unaware of the consequences of the rejection of these two suggested analyses of propositions about the experience of other selves, namely that the phenomenalist analysis breaks down unless Berkeley's problems are shown to be merely muddles, and a satisfactory phenomenalist account of other selves is provided. Since this is not a paper on phenomenism or on the nature of meaning, though I fear it may have seemed so too much already, we cannot here discuss these questions.

The consequence of all this so far as utilitarianism is concerned is, firstly, that the proposition 'Smith is pleased' is intelligible without entailing a behaviourist analysis; secondly, that this entails that whatever assertion can significantly be made by me about myself can *mutatis mutandis* equally significantly be made about others: the logical structure of the two sets of propositions is not radically different. And this brings me to my second and third questions: What do I mean when I assert that Smith is happier than I am? – which can be answered only if I know what I mean when I say that anybody is now happier (*a*) than himself at some other date, and (*b*) than someone else at some other or the same date.

UTILITARIANISM

III

Now it seems true to say that all words which grammatically possess comparatives and superlatives are the kinds of words which symbolise relations which in some sense are relations of magnitude, and therefore, it is sometimes supposed, make definite measurement possible. This is certainly the case with so called extensive magnitudes, where relations of magnitude can be expressed in terms of any arbitrarily chosen equal units. To say that X is larger than Y, or higher, or broader, is to make some assertion about the number of units common to some property possessed by both X and Y, such that the residue by which X exceeds Y is itself expressible in terms of such units, and is itself a property which could belong to some third entity Z, so that 'X is greater than Y' would be more precisely stated as 'X = Y + Z'. But it has always been recognised that such expressions as 'more than' or 'greater than' could also legitimately be used of relations different from those described above, different in at least one vital respect, namely that although the relations were still in some sense relations between entities of different magnitudes, yet the difference between the magnitudes was in no sense itself a magnitude; these relations hold between so-called intensive magnitudes, which are for this reason frequently compared with the relation which hold between qualities, where it is obvious that although qualities differ from each other in identifiable ways, the difference between two qualities is not itself a quality. The difference between a length of ten feet and one of eight is a length of two feet, but the difference between red and green is not a colour at all.

Now the difference of intensive magnitudes seem to occupy a curious intermediate position between these two great types of difference: whereas, as in the case of differences of quality, the difference between two intensive magnitudes is not itself an intensive magnitude – the difference between X, which is hot, and Y, which is hotter, is not a heat – yet it is a magnitude in that the entities possessing intensive magnitudes with respect to the same homogeneous qualities can be arranged in an objective graduated serial order, as qualities can not always be arranged.

This seems an obvious enough point, but those very philosophers who have made it best have usually not perceived its

UTILITARIANISM

equally obvious implications. If I say that X is hotter than Y and Y than Z, the hotness of X or Y or Z, by which I mean the *sense-perceived tactile sensation*, cannot be stated as a multiple of anything: it is nonsense to ask of anything whether its heat is the heat which is the difference between two other unequally hot objects. $X = Y + Z$ is obviously inapplicable here. What I can do, and what scientists in fact do, is to discover some reliably invariant correlation between these same data and something else – vibrations, dial readings, distances in the spectrum, anything which, being itself extensive, can be, in principle, measured precisely. Whether there be in reality any such primary qualities, in the seventeenth–eighteenth century sense of that term, or whether they are simply predictive formulae stating and prophesying correlations between sets of either sense-given qualities or of intensive magnitudes on the one hand, and of sense-given extensive magnitudes, that is, quantities proper, on the other, it is not our task to enquire. Whatever the answer, when we say of anything that it is an intensive quantity we mean to assert both that it has the relation of being greater or smaller, more or less, to something else, and that this is a transitive asymmetrical relation, but the relation, not of addable parts of wholes to each other, but of being arrangeable in some sort of serial order, correlatable with the number series, such that the general character of the serial order is itself directly given in experience.

It seems quite clear that pleasure, in so far as it qualifies states of mind, and is capable of having such relational predicates as ‘more or less’ significantly applied to it, is an instance of intensive magnitude. It seems clear that when I say of a given course that it is likely to give me more pleasure than another, or even only that I prefer to take it, I am therein asserting that I am acquainted with the fact, even if it be a fact only about expectation-data, that these data do in fact stand in a relation of being more or less something-or-other – being more or less pleasant-seeming, or desired, or approved of – than certain other data of the same species. That I am acquainted with complexes whose elements are connected by such relations and that the propositions affirming the relations are therefore synthetic, even if they be a priori, seems to me indubitable. It is only thus that I am able to construct a calculus of preferences, even if this calculus be one of relation with the facts of the physical world, or likely to be abandoned by me at the first opportunity, and so useless for practical life. All I maintain is that I

UTILITARIANISM

can say 'I should like a glass of wine more than a glass of water' and translate this into 'I believe that a glass of wine will give me more pleasure than a glass of water' and not mean the unmeanable – that is, that the difference between the pleasure expected to arise from the wine and that from the water is itself a pleasure which I could derive from some third entity. So that even if somebody does say 'I prefer a glass of wine to a glass of water but I do not prefer one glass of wine to two glasses of water', he does *not* mean that the pleasure of the first glass of water and the pleasure of the second glass of water are together greater than or equal to the pleasure of one glass of wine; for this would lead to the nonsensical proposition, 'Difference of pleasure is a pleasure': what he means is that the pleasure believed to attach to drinking one glass of water and then after it another, treated as a single complex or pattern of events, comes higher in somebody's scale, however that is derived, than that which he expects from the events consisting in his drinking one glass of wine on one single occasion.

The fallacy of supposing that increments of pleasure can be analysed into so many unitary increments, *minima sensibilia* or the like, which the economist Edgeworth suggested as a measure, really arises from the ambiguity of words like 'greater', 'less' and particularly 'equal to', which are thought to apply to amounts or quantities whose parts are equal to one another (whatever that itself may mean), whereas in the case of intensive magnitudes 'greater' and 'less' refer to places, notches as it were, on the scale, which like all scales can be, for convenience, represented by a spatial image and therefore divided into parts, but which in itself is simply a scale or series properly expressible not by cardinal but by ordinal numbers.

If this is so, then it follows that: (1) I can say 'X now gives me more pleasure than Y did a moment ago', which is an empirical proposition that asserts a causal connection between the effects on my state of feeling of X and Y respectively⁶ and an indication of the places occupied by the two states in the series in which they are arrangeable with respect to the property of pleasantness; and (2) I can say 'X gives me more pleasure than to Smith', or 'more pleasure to Smith than to Jones', and mean not that if we were all

⁶ Should this not be 'a causal connection between X and Y respectively and my state of feeling'?

UTILITARIANISM

presented with instances of X, a greater quantity of something would occur in me than in either Smith or Jones, but that, if it makes sense to say that we all feel pleasure in the same sense of the word 'pleasure', then the pleasantness of my state comes higher in my scale than his state in Smith's scale, or [Smith's state in Smith's scale higher than] Jones's state in Jones's scale, which presupposes that there is a method, in principle, of correlating our scales so that we could all be said to be feeling pleasure equally intensely; and this could, I suppose, be tested by the existence of telepathic intercommunication between us which was recognised *as* telepathic, so that I could feel Smith's state as my own, knowing it to be qualitatively identical with his, and discover by introspection or memory whether or no the pleasure was intenser than the one I should normally feel if I obtained the feeling directly, and not, as it were, via Jones [sc. Smith?]. This would be a test of the truth of, but not the analysis of the meaning of, this proposition, unless we adopt a behaviourist view of meaning, which, as we have seen, does not adequately describe our experience. No doubt for methodological purposes, as in the oddly named science of econometrics, it may be useful to ignore the fact that we do mean something validly by comparing our feelings with those of others, since only strict behaviourist criteria may assure the necessary reliability and precision, and above all measurability, of our results. But because a technique is useful it does not follow that only its results can be significantly asserted: because we may not in principle be able to demonstrate the existence of other persons, it does not follow that we do not understand what the proposition to be demonstrated *means*; indeed we should not suspect it of indemonstrability if we did not [understand it] – there would have been nothing to suspect.

It is at this point in the argument that the fallacy emerges on which we held that utilitarianism, and indeed quantitative ethics of any kind, finally founders. We can meaningfully assert that my state X is more pleasant than state Y, or that Smith's state X₂ is pleasanter than my state X₁. But what I *cannot possibly* assert is that my state X is pleasanter or less pleasant than Smith's state X₁ plus Jones's state X₂, that is, that if three persons are, say, equally happy, any two between them can be said to be happier than the third, that is, that *more happiness or pleasure* may be produced by making two men happy than by making one man so.

UTILITARIANISM

This proposition does not need labouring: yet the admission of its truth robs the Greatest Happiness formula of such remnants of significance as may have seemed to remain after the previous analysis of it. I can arrange the qualities of the states of various persons' minds in a order by correlating them with as many single mathematical series; I can correlate the various series to a single common series; and if I believe that the proposition 'I prefer state X to state Y' either entails or is equivalent to the proposition 'X is pleasanter than Y' I can truly say that there is a sense in which an increase in the intensive magnitude of a certain quality A in X from A_1X to A_2X would make X equal to Y or greater than Y with respect to the quality A common to X and Y, where 'greater than' and 'equal to' would refer to the relative positions occupied by X and Y in a scale or series; but I cannot significantly deduce from this than there is an increment symbolised as $A_2X - A_1X$ which, even if it cannot itself qualify anything, yet is abstractable in thought and therefore discussable and manipulable, that is, addable and subtractable, for that is what being an increment means. If that *is* what increment means, then another word must be chosen for differences of intensive magnitudes which makes clear that, although not identical with, it is more akin to, difference of specific quality in the same disjunctive set than to difference of parts of a conglomeration. And yet this proposition, which seems so self-evident on inspection, has been denied by philosophers who recognise the truth of the premisses from which it analytically follows. My clearest examples come from the works of McTaggart and Ross, and a more dubious one from those of Moore.

Thus McTaggart, in his *Studies in Hegelian Cosmology*, after giving the best and clearest account in English of this problem known to me, observes, in a passage doubtless familiar to everyone here, on p. 116, section 122:

If we have two pleasures of different intensities it is true, no doubt, that the excess of A over B is not *a* pleasure. For we cannot imagine that part of the intensity of A existing by itself. Its meaning depends on its being in combination with the rest of A's intensity. It would be meaningless to ask what the heat of an average June day would be like after the heat of an average December day had been subtracted from it. The remainder would cease to be what it had been, as soon as it was separated from the other part.

UTILITARIANISM

McTaggart's meaning is clear, and so far it is identical with the point I have used: but the expressions '*part of the intensity*' and again '*remainder separated from the other part*' are suspiciously ambiguous. In the next sentence he duly pays for this. It runs as follows:

But although the excess of A's intensity over B is not *a* pleasure, I submit that it is nevertheless, *pleasure*. Whatever has quantity must be homogeneous in respect of some quality, and is only quantitative in respect of that homogeneous quality. If therefore pleasure has an intensive quantity, then [this is the fatal step] each part of that quantity must be pleasure, including that part of which it is greater than another.

If then excess of A over B is pleasure, and a quantity, it must be capable of being brought into numerical relations with other quantities of pleasure. And thus, while it is true that we cannot imagine that excess as a separate pleasure, we can imagine a separate pleasure which shall be equal to that excess. If this is called C, then we shall be able to say that the pleasure in A is equal to the pleasure in B and C.

He goes on to say that he finds no difficulty in making such statements, and indeed has no hesitation in affirming that the pleasure he gets from a plate of turtle soup is more than twice the pleasure he gets from a plate of pea soup, or that the pleasure got from reading a new novel together with the pain of a hot walk to get it leaves a balance of pleasure greater than the pleasure from reading an old novel off his shelves; one may make errors in estimating the precise quantity of the increment, but one is liable to error in any such judgement; in principle we can and do make such judgements.

The fallacy ought by now to be fairly plain. McTaggart speaks of *parts* of intensive quantities, and of equality between them as implying equality of their parts. This is axiomatic of, because part of the definition of, extensive quantities; but is meaningless if applied to intensities, a word employed specifically in order *not* to imply that the quantities in question have *parts*. 'Equality' as a term is equally ambiguous: it entails divisibility in the case of *extensities*, but means something quite different, namely coincidence in a place in an ordered series, when used of intensities. McTaggart's

UTILITARIANISM

really gross confusion is evidenced by his two examples: the first both looks *and is* paradoxical, and is meaningless when analysed; the second seems much more plausible, but that is because it is not really analogous. To say that turtle soup is more than twice as pleasurable as pea soup is, if it is literal, strictly nonsensical; and is so because the notion of a residue is inconceivable. To say that I would rather have the new novel and the walk rather than the old is, of course, intelligible, because I am comparing two sets of possible acts: each is regarded globally as a single whole; and they are then assigned by me to their ordinal positions in my calculus. I am *not* adding the pleasure from the new book to the pain of the walk: I am mentally putting together the imaginary experience of reading the book and the imaginary experience of walking to get it in a imagined time series, and considering how pleasant the state obtained as a result and at the end of these would be relatively to another imagined state. I am comparing the pleasantness of A + B with that of C, not the pleasantness of A + the pleasantness of B with the pleasantness of C. The non-parallelism of the two cases is glaringly evident. The fallacy consists in regarding what he himself refers to as the quantity in respect of a homogeneous quality as always definable as a collection of detachable parts. He denies this in one sentence only to reassert it in the next. His account as it stands is instructively self-contradictory.

Ross in his *The Right and The Good* repeats this even more bluntly. On p. 143 of that work, after saying that pleasures have the only characteristic that is necessary to comparability, viz. difference of intensity, he goes on thus:

If this is clear we may next argue that pleasures are commensurable i.e. that one pleasure may not only be more intense than another but may be, say, just twice as intense. This surely follows from our previous conclusion. For if one pleasure is more intense than another, it must have a definite amount of extra intensity. True (as is often pointed out), the difference between two pleasures is not a pleasure, as the difference between two lengths is a length. This is part of what we mean by distinguishing intensity from extension. But the difference between the intensity of two pleasures is an amount of intensity, and of course a definite amount. And this may be exactly equal to the intensity of the less pleasure; and if so the greater pleasure is just twice as great as the less.

UTILITARIANISM

What follows is nothing of the kind. This passage seems to me to be almost self-refuting: for what is '*an amount of intensity*'? 'Pleasure', we say, is the name of the quality of an introspectible state; but what is an *intensity*? Surely the name of a particular relational property, of having a particular degree, that is, a place in a graduated scale. Relational properties cannot occur *in amounts*: to say this is to rob words of meaning. All that 'amount of intensity' can be in this passage is an elliptical way of saying 'amount of intensity of pleasure', that is, either 'degree of intensity of pleasure', which is not an amount of any sort, and cannot be twice or any multiple of anything; or else 'amount of pleasure' simply, that is, a pleasure of a certain magnitude, which Ross says *cannot* be the difference between pleasures. One degree cannot be twice another, even if the spatial distance covered by the pointer which registers it can. The contents of one beaker of water heated to the temperature of 40 per cent of a scale has not a degree of heat which equals the sum of the temperatures reached by the contents of two other beakers, each of which contains water heated to 20 per cent. 'Few people', says Ross on p. 144 of the same book, 'except those who are committed to hedonism would hesitate to say that an action of sublime heroism is many times as good a thing as some faint and passing shade of pleasure'; but if we are right, fewer still, be they never so much committed to hedonism, would on reflection hesitate to say that the above is not indeed false but meaningless. All that really follows from Ross's premisses, which seem quite correct, is that pleasures are somehow commensurable: but not in the way he indicates. If we enquire how this fallacy arises, the most plausible explanation is that it is due to the confusion of the behaviour of measuring rods or gauges with the character of that which is being measured.

The general nature of measurement is a difficult topic and I am not competent to embark on it. It will be sufficient to say that in order to obtain accurate results, that is, results expressible mathematically, we must correlate differences of quality with those of quantity. But correlation is not identification: the translation from one sphere into the other is a methodological desideratum, even if that is not all it is. Nobody now, except the most illiterate behaviourist writers, confounds what we mean by the word 'red' with what we mean by that number of light vibrations with which it is as a phenomenon correlatable. There is therefore no reference

UTILITARIANISM

to amounts of red, although we may speak of the amount of space coloured red, or of the degree of say, saturation of a specific hue of red. If a hedonometer were constructed – and there is at least in principle no reason why one should not be made – and if it correlates intensity of pleasure with some physical phenomenon, say the degree of blood pressure or the like, an amount which could be twice another amount would be the distance traversed by the pointer over the dial of the hedonometer. And as both the size of the unit and the zero-point might vary with types of instrument used, the so-called increment of one pleasure over another could be stated only by reference to the type of instrument used, and though itself as a datum absolute in magnitude, could be described only in terms of the readings of a particular type of hedonometer, which could be generalised by fixed transformation rules for translating this into readings of the other types. The quantitative statements, though symbolising something, could not symbolise amounts of the quality of felt or anticipated pleasure: the supposition that they would or could, which the amount-theory entails, is what vitiates the arguments of Ross and of McTaggart.

My third instance come from Professor Moore's *Principia Ethica* and refers not to pleasure but to what he there refers to as the goodness or the value of something. This is relevant because I take it that, whatever goodness or valuableness may be, if it possesses a dimension at all it is of the order of intensive and not extensive magnitude: to say that goodness has parts seems even more extraordinary than to say that pleasure has. Yet Moore appears to think not only of goods, but of goodness, as addable and subtractable. Thus on p. 28 of the book, in discussing intrinsic value in relation to what he proposes to call the principle of organic wholes, he says:

It is certain that a good thing may exist in such a relation to another good thing that the value of the whole thus formed is immensely greater than the sum of the values of the two good things. It is certain that a whole formed of a good thing and an indifferent thing may have immensely greater value than that good thing itself possesses. [And a few lines below:] Whether the addition of a bad thing to a good whole may increase the positive value of the whole or the addition of a bad thing to a bad whole may produce a whole having positive value may seem more doubtful: but it is at least possible. However we may

UTILITARIANISM

decide particular questions the principle is clear. The value of a whole must not be assumed to be the same as the sum of the values of its parts.

It may seem very strange to accuse Professor Moore of all philosophers of the sin of ambiguity, and yet this is what we find we are compelled to do. For his last summarising statement may mean two entirely different things: either it means that the combination of two elements which in isolation possess certain values such that each occurs at a certain point of our valuation scale (which may or may not be subjective) itself as a complex possesses a value different from the value which either element would produce if combined not with the other, but with some *third* entity whose value in isolation is held to be equal to the value of the one which did in fact enter the complex. That is to say that the complex AB may have a value different from the value which would be possessed by a complex AC or AD, where C or D are each precisely as valuable as B; and it may in passing be noted that AC and AD need not even be complexes constituting a real whole, like AB, but be simply sets of externally related events. All that then follows is that knowledge of the values we ascribe to particular entities in isolation does not enable us to deduce the value which would belong to the complexes of which the entities are constituents. If this is what Moore means then he is very likely right, and the principle is true and important. But his proposition ‘The value of a whole must not be assumed to be the same as the sum of the values of its parts’ may also mean – indeed would naturally be taken to mean – something quite different, viz. that *values* themselves can be added into sums, that is, that not only goods have parts, but their goodnesses: that is, that we could actively add the *values* – whatever these may be analysed as, whether as qualities, relational properties or relations – and by this process we should, so it seems to be asserted, get a certain sum total, only, as a matter of fact, we should get the sum wrong if we did so, or rather the sum we should get would not be the answer to our question, because some other sum is. This entails the proposition that not only valuables but values are addables, but the result obtained by adding valuables cannot be translated into the sum obtained by adding the corresponding values. Values as addabilia are simply Ross’s hypostatized intensities once again, and the fallacy on which the proposition rests is Ross’s fallacy. To add

UTILITARIANISM

the goodness of A to the goodness of B, whatever the word 'goodness' be analysed to mean, is an impossible process. The proposition that it would lead to this or that result is strictly nonsensical, since there is nothing of which it would conceivably be true or false.

I do not know if Professor Moore would agree to even the first interpretation of his statement, but it is the second which, if not he himself, yet many others have undoubtedly held – both taken his general thesis to imply, and have themselves held – with regard to value at large or pleasure in particular. If we are right in maintaining this view to be founded on a fallacy, then it follows that, since pleasure or goodness or happiness is not quantitative in the sense of having parts, we cannot significantly say in answer to our Fourth Question that if $A > B > C$, then $A + B$ is a fortiori greater than C. But if this is so, then to say that a given course of action, in making more people happy than any other, *may* produce a greater sum of happiness than another is meaningless, for to say that if A is as happy as B and B as C, then the happiness of $A + B$ is a greater sum of happiness than the happiness of C alone, is meaningless. For it is as if we were asked to compare the length of one building with the height which two other buildings possessed between them, without being allowed to conceive of them as one on top of the other, but only of their heights as being so: which means nothing at all. And this follows, whatever intensive magnitude we like to take: whether desires, or preferences, or approvals, or any state which can be intenser than another. It therefore cuts across the issue between naturalism and apriorism, subjectivism and objectivism, and applies equally whichever of these views we adopt. The expression 'greater happiness' or 'greater happiness than' can hold only either of the relations of one of my states or complexes of states to another belonging to me, or to the relation of my state to that of someone else. The relation in both cases is, if the ostensive meaning of the terms remains unaltered, transitive and asymmetrical. The terms are not terms *of a* before and after series, for if they were, no state could ever be said to *equal* another in intensity, since each term of the series is either before or after any other term, by definition. The particular psychical states about which the propositions are made are as it were *on* a series, correlated to it, in such a manner that absolute equality means absolute coincidence, absolute qualitative identity in respect of the relevant quality, and numerical difference of the

UTILITARIANISM

particular states. In this respect intensive quantities are precisely analogous to qualities; and the same serial or dimensional order characterises them as that which characterises the members of certain disjunctive sets; the fundamental relation being similarity. The similarity I mean is resemblance proper, and not partial identity. The sense in which a specific shade of yellow and a specific shade of green are similar and given as such is not the derivative sense in which two patches coloured with an identical hue are said to resemble each other in both being coloured, that is, in being determinants of the same determinable.

Digression

Here I propose to introduce a short digression on the nature of propositions affirming such qualitative similarity, which is relevant is so far as the relation between one intensity of, say, pleasure and another seems to me to be fundamentally an ordinal relation between the terms of an infinite series, and to be therefore to a certain extent classifiable with ordinary qualitative series. To make this clearer: the question is, of what are we aware when we grade classes of pleasure in order of intensity? It seems to me that when I say yellow is more like green than blue,⁷ where yellow, green and blue are not determinables but the determinate shades such as qualify particular sense data, those philosophers are right who say that this proposition is neither inductive nor analytic in any ordinary sense of the word. It is not inductive because further experiences can make no difference to the relation detected by me as holding between these three – if I may borrow an American term – given colour qualia which are, like Locke's simple ideas, not particulars, in that they are indefinitely repeatable, and not class-names, in that they are not determinables but ultimate determinates.

Those, consequently, who regard all propositions as either empirical or analytic are forced to say that, if my proposition is not about empirical particulars at particular dates and places, but genuinely general in intension, I *mean* by 'more like' certain relations of proximity, if not of bands in the spectrum or the frequency table of vibrations, then at least of somebody's

⁷ This issue comes up in IB's 1937 published article, 'Induction and Hypothesis', which is why I tentatively assign the date c.1937 to the present paper.

UTILITARIANISM

particular colour atlas. If a certain shade of yellow is called No 17, a certain shade of green No 24, and of blue No 28, 'more like' merely reports this verifiable fact. It merely tells me how, if I do not know what these words mean in a language, I can discover their meaning by being referred to a number and looking it up in the atlas. To say, therefore, of the yellow quale that it is like the green quale in a certain specific way is no more than to say that it is the yellow it is, and the green the green it is. The relation is what Hume sometimes calls philosophical, a relation between words or symbols or other catalogue entries in the colour atlas.

But, I may ask, how is the colour atlas itself constructed? It may be constructed for some specific scientific purpose; the order of colours may be regulated by a one-to-one correspondence with the order of the relevant light waves. In which case the order will not be an order of graduated resemblance, that is, not the sort of colour atlas I refer to, if indeed I refer to one at all; and this because the only colour atlas I can be referring to must be such that in it the colours are arranged not arbitrarily but in a order dictated by something in themselves, that is, in some sense it must be a *map* of what I find in sensible reality. The relation between the numerical values assigned to the hues of such an atlas is not arbitrary but made to correspond with a real relation of greater or lesser similarity between qualia. The numbers record resemblance, which cannot to be defined in terms of them. In which case I can eliminate the map, and directly intuit the relation between the qualities of sense data – a relation which may best be described as the one which holds between the ribs of a coloured fan where the colours insensibly pass one into the other in a continuous series, such that were a gap to occur I could fill it in, as Hume observes, for the imagination, without having to have seen it in the external world. And judgements founded on such empirical intuition will be derived from experience, be a priori, and, as it seems to me, synthetic.

I am far from clear or happy about this, but it seems to me evident that we are continually making such intuitively certain judgements about the characteristics of the sensible world, and could not be acquainted with anything at all if we did not. The method is usually called intuitive induction, and seems to me to precisely the same as that employed in comparing or verifying intensities of pleasure or of pain. The qualia which constitute the pleasure–pain series have no specific names, but neither have the

UTILITARIANISM

absolutely determinate visual or tactile qualia with which we are acquainted and which we recognise by direct acquaintance as repeated in each new manifestation. If Hume were right about resemblance as a philosophical, that is, verbal, relation it is difficult to see how any synthetic proposition could even be constructed. All comparison, abstraction, recollection presupposes acquaintance with this fundamental category – called by Carnap, I think, *Teilähnlichkeitserinnerung* – which characterises the primary data from which we start. It is this faculty of intuiting the difference of structure between intensity and extension that enables me to assert that pleasure belongs to one category rather than the other. This assertion itself seems to me neither analytic in the sense that it follows from arbitrary definitions, not empirical in the sense that further experience can modify it or have any bearing whatever on it.

If this is so, if, that is, the question ‘Is X happier or better or more intensely desired or approved of than Y and Z?’ is a meaningless question because the qualities are not addable, nor are they intensive quantities, and to speak of sums or increments or residues in this connection is meaningless, what becomes of the greatest happiness principle? Or of any ethical theory, for that matter, which entails that characteristics are addable in any fashion? Unfortunately we cannot leave the matter like that. When we say that more general happiness will be created if we spend our money on supporting scientific research or rescuing victims of persecution or giving presents to our friends than by putting it in stockings or presenting millionaires with cheques for £5, or even by making a thousand people happier than one, we may be speaking loosely, and philosophers may be misled by our usage, but we undoubtedly mean something, and it is important, if we are not to be misled further, to know what it is that we do mean. If we substitute for happiness the even vaguer concept of welfare we are enabled to speak of social progress: and this looks like entailing the commensurability of whole clumps of characteristics each of which belongs to some particular individual person. And although such concepts as welfare or progress are very unclear, and may even on a pessimistic view be illusory in the sense that they have no application in human history, we are not prepared to consider them meaningless or self-contradictory. But if such expressions meant literally what they said, we should have to begin a Kantian

UTILITARIANISM

deduction of the category of the amount of happiness or of welfare to see what else must be presupposed in order that these curious types of fact should exist, as we should hold they did exist. But no category of the meaningless is possible, and so some interpretation other than the literal one must be sought.

I confess I do not know what the right answer is; but I provisionally suggest the following. Whenever we speak of whole groups or societies as being happier or more satisfied or more progressive than they were, or than some other society is or was, we do not mean that in one there are more individuals who are each above zero point (irrespective of distance from zero) in respect of the characteristic which we are ascribing in unequal degrees to the two societies; for we should not naturally say that a community of a hundred of which all the members are just happy enough to continue to live rather than commit suicide was happier than one of fifty whose members each led reasonably contented and interesting lives. What we should wish to say is, I think, that the happier society is so called in virtue of the fact that an average member of it is, between the dates of which the proposition holds, happier or more satisfied than he either was or would have been before in the same society, which, *ex hypothesi*, used to be less happy in the past, or else would have been in some other past or present society. And the relative intensity of happiness, whether momentary or over a length of time, is verifiable by memory or by introspection or, in ideal cases, by telepathy, as the case may be. I am not quite sure how to analyse the descriptive phrase 'average member'; the one analysis which may be ruled out as nonsensical is in terms of total happiness divided by number of persons. The sort of analysis wanted is, I think, into a fictitious type-character constructed out of the largest number of characteristics common to the actual members and lacking in any characteristics incompatible with these.

If we say that we ought to promote, or that it would be good to promote, greater social happiness, and that this act will do so best, we do not mean that more wants will be satisfied by it regardless of their individual intensities, each want to count for one and for no more than one, although the virtual unverifiability of intensity may lead to the disregard of intensity in practice by administrators or social scientists. What we mean is that as a result of the act the community would possess a certain quasi-dispositional characteristic in a greater degree, the characteristic being one

UTILITARIANISM

which we think good, or approve, or approve of approving, in this case happiness. This it certainly could not be said to possess unless the average individual in the society would be made happier by this than by any other arrangement. This at least is entailed; but possibly we mean more than this. Possibly we also may wish to denote a collective, gestalt quality which societies possess, such as harmoniousness or stability, which qualifies social wholes but not necessarily any parts of them. Needless to say we need mean no more by 'wholes' than complex hierarchies of logical constructions out of the events which constitute the lives of the individual members of the society, and such gestalt qualities, if there be any, as quantify any conjunction of them.

I am not at all sure that this is what we do mean, and feel particularly uncertain about the concept of the average individual, which seems central to this thesis. But if even approximately correct it does at any rate enable us to grade whole communities in scale of preference, not irrationally but in virtue of specifiable characteristics of which some qualify societies only in the sense that they qualify individual men and women, and some qualify the wholes themselves in a more direct fashion. I feel particularly doubtful about this latter point; but even if it be ignored, verification and induction become possible, and utilitarianism, whether ideal or hedonistic, thus clarified, whether its central proposition is analytic or synthetic, becomes tenable.

This result will disappoint only those who hoped to find in utilitarianism a theory which permitted, at least in principle, verification by exact measurement; of which only extensive magnitudes are capable. Such a hope rests on a false belief; it is no kind of answer because the question is an improper question; and as long as it goes on being asked among ethical philosophers, inevitable confusions will vitiate their enquiries.

In conclusion I should like to say that I feel far more certain of the absurdity of the view which I attack and of the need to attack it than of the correctness of the solution which I end by offering. It will seem very tentative and unexciting by comparison with the more forcible solutions of such an outstanding problem as the proper method of ethics. But there is no a priori reason why we should expect the truth, when found, to be dramatic:⁸ and it

⁸ This version of C. I. Lewis's dictum may give a *terminus ante quem* of c.1937 for its original utterance.

UTILITARIANISM

seemed worth while to attempt to clear away certain muddles and confusions which to this day continue to make moral philosophy needlessly difficult and obscure.

© *The Isaiah Berlin Literary Trust* 2002

Posted 16 May 2002