



A Reconstruction of Classical Utilitarianism

Tom Warke ¹

How to cite: Warke, T. 'A Reconstruction of Classical Utilitarianism.' *Journal of Bentham Studies*, 2000, 3(1): 3, pp. 1–20. DOI: <https://doi.org/10.14324/111.2045-757X.014>.

Published: 01 January 2000

Peer Review:

This article has been peer reviewed through the journal's standard double blind peer review.

Copyright:

© 2000, The Author(s). This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC-BY) 2.0 <https://creativecommons.org/licenses/by/2.0/>, which permits re-use, distribution and reproduction in any medium, provided the original author and source are credited • DOI: <https://doi.org/10.14324/111.2045-757X.014>

Open Access:

Journal of Bentham Studies is a peer-reviewed open access journal.

¹ University of Sheffield, UK

A Reconstruction of Classical Utilitarianism

TOM WARKE

Economics, University of Sheffield

1. Introduction

Although it is unusual for an author to describe his or her personal odyssey toward a particular paper, I judge that a breach of academic protocol can be justified in this instance by its utility for a multi-disciplinary audience. My interest in classical utilitarianism derived from discomfort with the modern use of the utility concept in my own discipline of economics. This discomfort arose not from any doubts about the logical force of economic analysis, but rather from doubts about its presumption that individual optimal choice in practice is determinate. Determinate optima follow from the two axioms, rarely questioned by economists, that agents' preference sets are complete and transitive. Transitivity is the logical constraint on each agent that all outcomes which are preferred to an arbitrary outcome A must also be preferred to all outcomes that are not preferred to outcome A. Completeness is the claim that an agent faced with any two outcomes A and B must have one of three mutually exclusive responses: A strictly preferred; B strictly preferred; or strict indifference between A and B. Taken together, these two axioms generate the utility concept of modern economics - a one-dimensional determinate ordering over every outcome that an individual agent might possibly contemplate. The heresy that I had long harbored was that utility is in reality multi-dimensional, and that the term in economics names an *index number* ordering of outcomes, each of which encompasses an array of utility components that are less than perfectly commensurable to the agent. Subject, then, to the index number problem, the agent's ordering is inherently ambiguous: faced with two outcomes A and B, the agent might have a *continuum* of responses between A strictly preferred and B strictly preferred. In short, preference sets are fuzzy sets and agents often dither. With preference set boundaries undefined, optimality is problematic and transitivity in choice need not occur.

My search for the historical antecedents of utility in economics led me to W. S. Jevons, and from Jevons to Jeremy Bentham, in whose works I found a utility concept

essentially identical to my own.¹ From Bentham I moved to J. S. Mill, with an initial interest in Mill's degree of fidelity to the utilitarianism of his childhood mentor. My reading of *Utilitarianism* rated it substantially Benthamite, with useful extensions in the cases of acquired tastes and habitual behavior. It was only at that point that I began to read the philosopher critics of Mill, beginning with G. E. Moore in *Principia Ethica*. My immediate reaction, which has been reinforced over time, was that Moore had simply failed to comprehend classical utilitarianism, and that this rendered his critique useless and his continuing influence most regrettably mischievous.²

My purpose in this paper, therefore, is to set out classical utilitarianism as I understand it. Because I shall find it convenient to use mathematical concepts that were unavailable to Bentham and Mill, I shall try to be scrupulous in indicating where mathematics imposes more rigor than their original system contained. As many readers will be more knowledgeable in the wider arena of ethical philosophy than I am at present, they may find helpful an immediate listing of my main referents for this paper. My understanding of modern utilitarianism is based upon Hare, Harsanyi, Quinton, Goodin, and Scarre.³ My comments on Moore (whom I do not regard as a utilitarian, ideal or otherwise) derive from my own reading of *Principia Ethica*, but I have otherwise relied mainly upon the superb exposition in Warnock as a secondary source for non-utilitarian ethics in the first half of the twentieth century.⁴ And finally, on the advice of John Deigh, I have taken Berger as the definitive contemporary interpretation of Mill's *Utilitarianism*.⁵

2. Bentham's Conception of Utility.

At several points in *Utilitarian Ethics* (1989), Anthony Quinton criticizes as 'oversimplified' Bentham's 'definition of happiness as a sum of pleasures'.⁶ The criticism

¹ The process by which economists turned the utility concept into a one-dimensional determinate ordering is described in T.W. Warke, 'Mathematical Fitness in the Evolution of the Utility Concept from Bentham to Jevons to Marshall', *Journal of the History of Economic Thought*, xxii. (2000), pp. 3-23.

² See G.E. Moore, *Principia Ethica*, Cambridge, 1903. A more detailed presentation of my case against Moore is given in 'Multi-Dimensional Utility and the Index Number Problem: Jeremy Bentham, J. S. Mill and Qualitative Hedonism,' *Utilitas*, xii. (2000), pp. 176-203

³ R.M. Hare, 'Ethical Theory and Utilitarianism' in *Utilitarianism and Beyond*, ed. A. Sen and B. Williams, Cambridge, 1982, pp. 23-38; J.C. Harsanyi in *ibid.*, pp. 39-63; A. Quinton, *Utilitarian Ethics*, 2nd edition, London, 1989; R.E. Goodin, 'Utility and the Good' in *A Companion to Ethics*, ed. P. Singer, Oxford, 1993, pp. 241-48; and G. Scarre, *Utilitarianism*, London, 1996.

⁴ M. Warnock, *Ethics since 1900*, 2nd edition, London, 1966.

⁵ F.R. Berger, *Happiness, Justice and Freedom*, Berkeley, 1984.

⁶ Quinton, *Utilitarian Ethics*, p. 45.

would be justified were it accurate, but in fact it is abundantly clear throughout his writings that Bentham regarded happiness, or utility, as a *vector* of pleasures (and freedoms from pain). A vector maps to a sum only if its elements are equally weighted, and Bentham's method for estimating the value of a pleasure or pain for an individual agent shows that he imposed no such condition. In particular, Bentham distinguished between what I shall term the *dose* of a potential pleasure or pain - its duration discounted for remoteness and uncertainty - and the *intensity* with which a person responds to that specific stimulus. For Bentham, the happiness associated with a given act was determined jointly by the act's pleasure/pain doses and by the intensity weights of all interested agents. I would further argue that Mill was showing the same perspective, with 'quantity' referring to dose and 'quality' referring to intensity of response, when he stated in *Utilitarianism*: 'It would be absurd that while, in estimating all other things, quality is considered as well as quantity, the estimation of pleasures should be supposed to depend on quantity alone'.⁷

For Bentham and Mill, the variability of intensity weights, both across different types of pleasure and pain for any given agent and across different agents for any given type of pleasure or pain, meant that neither individual nor aggregate happiness could be expressed as a simple sum of pleasures. Furthermore, Bentham recognized that an inherent imprecision of intensity weights carries over into imprecise happiness comparisons among alternative choices. A net pleasure advantage between some pairs of acts will be obvious to an agent, but for other pairs the optimal choice will be problematic even if their pleasure/pain doses are fully known. Though he could have known neither the term nor its formal properties, Bentham treated the utility associated with 'complex exciting causes' (that is, any compound mixture of pleasures and pains) as an index number.

In order to illustrate Bentham's method we must begin, as he did, with the measure of a single pleasure or pain to a single agent:

To a person considered *by himself*, the value of a pleasure or pain considered *by itself*, will be greater or less, according to the four following circumstances:

⁷ J.S. Mill, 'Utilitarianism' in *The Collected Works of John Stuart Mill*, ed. J.M. Robson, Toronto, 1969,

1. Its *intensity*.
2. Its *duration*.
3. Its *certainty* or *uncertainty*.
4. Its *propinquity* or *remoteness*.⁸

Bentham's most detailed instructions for consolidating these circumstances appeared in the *Codification Proposal* (1822), where he singled out intensity as 'not susceptible of precise expression' while claiming that the 'others *are*'.⁹ To illustrate Bentham's method, let us consider as our complex exciting cause a proposed visit by two friends, Geri and Ben, to an exhibit of Robert Mapplethorpe photographs.

Bentham's first step in evaluating an act calls for a listing of every potential pleasure or pain associated with it. These are, we may suppose, pleasures of the eye and of amity; and possible pain of impiety (I refer, of course, to some of the subject matter of Mapplethorpe's work and its condemnation by certain religious authorities). Bentham presumed in the *Codification Proposal* that the duration of each pleasure and pain could be specified, together with a probability that the anticipated sensation will in fact be realized, thus yielding the duration's expected value.¹⁰ In addition, he applied an across-the-board discount factor for the remoteness in time of the proposed event, for which he used the prevailing money rate of interest on the grounds that it measures a generalized pain of postponement.¹¹ By this procedure, the dose of each pleasure and pain can be precisely expressed in cardinal units: discounted expected durations of time. Suppose, then, that our proposed act of attending the Mapplethorpe exhibit (Act M) yields the following dose vector:

x. pp. 205-59.

⁸ *An Introduction to the Principles of Morals and Legislation*, ed. J.H. Burns and H.L.A. Hart, Oxford, 1996 (*The Collected Works of Jeremy Bentham*), p. 38. Hereafter *IPML* (CW).

⁹ *'Legislator of the World': Writings on Codification, Law and Education*, ed. P. Schofield and J. Harris, Oxford, 1998 (CW), p.254.

¹⁰ Neither Bentham nor Mill paid much heed to the distinction between utility, a property of an object as perceived by an agent, and happiness, a property of the effect that the object has on the agent (see J. Broome, 'Utility', *Economics and Philosophy*, vii. (1991) pp. 1-12). Bentham's procedure relates validly the amount of potential hedonic pleasure from a choice under consideration to the amount of realized hedonic pleasure once the choice is taken only if expectations are accurately represented by an objective probability distribution that is known to the agent. In economists' terms, Bentham's hedonic calculus applied to future consequences of alternative choices carries an implicit assumption that agents have rational expectations.

¹¹ *Legislator of the World* (CW), p. 251. Bentham's use of an external discount rate to reflect the pain of postponement is required in order that doses be independent of agents. It would seem more realistic to presume that agents apply their own discount rates, and that these differ between types of pleasure and pain, but perhaps Bentham's procedure is justified by the usefulness of the distinction between objective dose and subjective intensity that he wanted to achieve.

Figure 1: Pleasures/Pain of: the eye amity impiety

ACT M	ACT A
12	6
6	6
8	0

By exactly the same procedure, we might calculate the doses of pleasures (with no pains in this case) from an alternative act of viewing a smaller exhibition of Ansel Adams photographs (Act A). Now, *if* happiness were a sum of pleasure quantities minus pain quantities, we would find that Act A outranks Act M (12 net pleasure units to 10), *independent of any agent*. With Bentham's felicific calculus, however, rank orderings are meaningless in the absence of interested persons with their respective intensity weights. This has two vital implications that have been largely ignored or misinterpreted by subsequent philosophers. The first point has already been made. Insofar as an agent is unable to assign precise relative weights to the disparate stimuli of beauty, amity and impiety, so too must there be imprecision in the rank ordering of acts that entail varying mixes of them. I have argued elsewhere that Bentham's recognition of this ambiguity is reflected in the very coarse-grained analyses to which he confined himself.¹² Bentham was adamant that the greatest happiness principle furnished the sole valid criterion for judgements of right and wrong, but he never claimed that it could always give a determinate fine-grained answer. In this paper, however, I wish to stress the second implication of Bentham's utility concept, for it profoundly affects interpretations of classical utilitarian ethics. This second point can be made more forcefully if we provisionally suppress the inherent fuzziness of intensity weights and allow our agents to assign them with precision.

Suppose, then, that we bring in Ben, whose aversion to impiety is shown by the intensity weights [0.3, 1.2, -0.5] on the respective types of pleasure and pain. The felicific calculus reveals, for Ben, a significant net pleasure advantage of Act A over Act M (9 net pleasure units to 6.8). Ben's friend Geri, though she shares his relative weighting of visual and amity pleasures, has a total disregard for impiety pain, shown

by her intensity weights [0.2, 0.8, 0]. As a result, Geri finds more happiness with Act M than with Act A (7.2 pleasure units to 6). The crucial point is that the ‘rightness’ or ‘wrongness’ of acts, given their doses of potential stimuli, depends entirely upon the tastes and preferences of interested agents. Moore’s procedure of discovering first what is ‘good’ and *then* examining whether the consequences of particular acts bear this intrinsic worthiness, independent of who is affected and how they respond, is totally irrelevant to the Bentham/Mill standard of conduct. In our particular example, Moore would presumably credit beauty and amity with ‘goodness’ and ignore impiety. As the doses stand, Moore could designate Act M unequivocally as the right choice for every agent (if we were to change Act A’s amity dose to anything between 6 and 12, however, Moore would face a weighting problem between his two goods that *Principia Ethica* does not tell us how to resolve).

In contrast to Moore, classical utilitarianism attributes no *intrinsic* worth - no Moorean ‘goodness’ - to act characteristics *per se*, for nothing has value in the absence of identifiable sentient beings to whom it gives pleasure. For example, there are no artistic creations whose beauty entitles them to exist for their own sake, as Moore would have it. According to classical utilitarianism, the value of visual art depends entirely upon the existence of someone, somewhere, at some time, who desires to see it (including, of course, the artists themselves). This position was stated with absolute clarity by Mill in a much-derided passage from *Utilitarianism*:

In like manner [to visible and audible], I apprehend, the sole evidence it is possible to produce that anything is desirable, is that people do actually desire it.¹³

In my view, the derision expressed by Mill’s critics, from Moore to Bertrand Russell to the present day, merely reflects the critics’ miscomprehension of the classical utilitarianism that Mill was faithfully expounding. Although critics are free to propose an *alternative* ethics that designates a priori what is worthy to be desired, Mill, within *his* system, committed no fallacy, naturalistic or otherwise, by taking ‘desirable’ as meaning ‘able to be desired’ and nothing more. By my interpretation, classical

¹² In addition to the places already cited, I have further argued this point in a paper not yet accepted for publication: ‘Classical Utilitarianism and the Methodology of Determinate Choice in Economics and in Ethics’. I am happy to supply copies to anyone interested.

utilitarianism grants unconstrained ethical autonomy to prudent agents making choices with zero extent (that is, choices affecting no other sentient being): what they choose to do is the right thing to do.

Returning to my example, suppose our two friends are not free at the same time during the exhibit period, so that amity pleasure is not available from either act. Ben would rightly choose Act A (and would rightly choose to stay home rather than visit the Mapplethorpe exhibit alone) and Geri would rightly choose Act M over Act A. If, on the other hand, they have simultaneous free time sufficient to visit just one exhibit, then their joint decision is no longer one of zero extent. In the latter case, designation of the right outcome requires a further weighting between their divergent interests, for it is no longer true that psychological hedonism and ethical hedonism necessarily coincide.

My next contention regarding Bentham's utility concept concerns his method of resolving conflict between psychological and ethical hedonism - his method of designating right conduct among acts with non-zero extent, affecting agents whose interests conflict due to their differing intensity responses to the relevant pleasures and pains. In my view, Bentham's greatest happiness principle as a means of rank ordering such acts is distinguished by its egalitarian interpersonal weighting rule: 'Everybody to tell for one, nobody for more than one.'¹⁴ Any different interpersonal weighting rule (as, for example, John Rawls would impose) generates a different ethics. Whether such an ethics could still be classified as utilitarian would depend, I believe, on its other features, as I shall discuss in Section Three. Also in Section Three I shall deal with the question of whether Mill adhered fully to egalitarian interpersonal weighting and, indeed, whether Bentham himself did so consistently.

The vector notation that I have adopted to illustrate my interpretation of Bentham's utility concept was not available to Bentham. Within it, however, his egalitarian rule is satisfied by a normalization of each individual's relative intensity weights such that they sum to a common number. This having already been done in my example (to the value +1), Bentham's greatest happiness principle can be applied simply by comparing the pleasure sums of the two acts across the two agents. We thus

¹³ Mill, 'Utilitarianism', p. 234.

¹⁴ Now that the original source of this maxim has been discovered by Philip Schofield, in the 1827 edition of Bentham's *Rationale of Judicial Evidence*, I would like to initiate the custom of using Bentham's word 'tell' rather than the word 'count' that appears instead in Mill's version, given that Mill meant to attribute the phrase to Bentham anyway.

determine that the right joint decision for Geri and Ben according to classical utilitarianism is Act A, for it yields 15 net pleasure units as compared to 14 for Act M.

As in any ethical system, however, the right decisions are not always made. Although Ben enjoys a harmony of interest with the ethically right choice, for Geri there is conflict. Since each of them has an interest in coming to agreement (for each of them, the other's first choice is more desirable than staying home or going alone), there exists an incentive to negotiate. But the actual outcome will depend upon their respective bargaining skills and strategies rather than the dictates of utilitarian ethics. Geri's disharmony of interests in my example thus raises the issue of utilitarian duty. Supposing that Geri is a utilitarian, does she have an *obligation* to follow the dictates of the felicific calculus?

I believe that Bentham's answer would be 'No', for I concur with H. L. A. Hart that Bentham regarded the greatest happiness principle as 'a standard of morality not for the *guidance* of individual conduct but only for the critical *evaluation* of it'.¹⁵ Geri is motivated to get her own way, just as Ben is motivated to get his own way. If she has the bargaining skill and status when they make their joint decision to override Ben's preference, then an observing Benthamite can observe that her dominance produced the wrong outcome but would not condemn her for moral laxity. Mill's answer would, no doubt, be more complex. I shall postpone to Section Four a discussion of Mill's treatment in *Utilitarianism* of the imperatives of virtue, but will note here that it represented a considerable retreat from his position in the 1830s, at the peak of his reaction against his Benthamite upbringing.

At that time Mill said, 'Nothing is more curious than the absence of recognition in any of [Bentham's] writings of the existence of conscience'.¹⁶ He went on to portray conscience as a powerful force for inducing ethical behavior. Presented with my example, Mill might have argued that Geri's conscience *should* lead her to the correct ethical choice, voluntarily bearing the smaller of the potential pleasure sacrifices (1.2 units rather than Ben's 2.2) and going along with Act A. But by the time Mill came to write *Utilitarianism*, this role of conscience had been superseded by a more sophisticated analysis. I suspect that Mill had by then realized that Bentham was absolutely right to ignore conscience as an independent ethical force in

¹⁵ *IPML (CW)*, pp. lxxix-cxii.

¹⁶ Mill, 'Utilitarianism', p. 95.

his system, for it serves only to vitiate utilitarianism's defining feature, namely its provision of an *external* standard for right and wrong.

Suppose, for example, that Ben's conscience tells him to cause no harm to Geri, despite the sacrifice to himself, so that he insists on Act M. If conscience operates independently of pleasure and pain, then we can no longer rely upon psychological hedonism to predict which outcome an agent will promote. More importantly so far as ethics is concerned, conscience furnishes no obvious criterion for designating an outcome as right. It would surely be ludicrous, for example, to attribute rightness to the greatest self-sacrifice independent of its other consequences. The greatest happiness principle under egalitarian interpersonal weighting, by contrast, is clearcut (within the precision range, we should remind ourselves, with which intensity responses can themselves be expressed by individuals). A Benthamite observer might *admire* a person whose conscience promulgates a self-sacrificing joint decision, but should that person in our example be Ben, the Benthamite would still designate the outcome as wrong.

For Bentham himself, the primary importance of a clearcut rank ordering among acts with non-zero extent was not to show individual agents what to do, but rather to show utilitarian legislators and judges what to do. To conclude this section, let us next suppose that the City Council where Ben and Geri live is debating whether to sponsor a Mapplethorpe exhibition or an Adams exhibition. Bentham's greatest happiness principle would designate as the right choice the one with the larger pleasure value, summed over all agents within the jurisdiction, using egalitarian interpersonal weighting. In order to secure this outcome, public choice decision-makers must be motivated to apply the utilitarian criterion (rather than indulge their own personal preferences), and they must then be able to estimate accurately what are the *mean* intensity weights on all relevant pleasures and pains among the constituents they serve. Glossing over these thorny political problems, my final point in this section is that the right outcome depends once again upon the composition of tastes and preferences among the affected agents. If, for example, everyone is either a 'Geri' or a 'Ben', then it is right for the City Council to sponsor the Adams exhibition if the relevant population consists of 36% or more Bens, but it is right to sponsor instead the Mapplethorpe exhibition if the share of Bens is 35% or less (recalling that the pleasure loss from visiting the less-preferred exhibit is 2.2 for Ben and 1.2 for Geri, the value of x at which $2.2x$ equals $1.2[1-x]$ is 0.353).

3. An Axiomatic Description of Classical Utilitarianism

In this section I intend to present my interpretation of classical utilitarian ethics more concisely, as a set of four axioms, and to consider briefly the consistency with which Bentham and Mill (of *Utilitarianism*) adhered to each of them.

The first axiom I would term Euclidian, in the sense that Bentham and Mill took it to be a self-evident proposition, the foundation of their entire system and unsusceptible to proof: **Axiom 1.** It is better for a sentient being to experience pleasure than pain. Bentham called this, simply, ‘the principle of utility’ and said of it:

Is it susceptible of any direct proof? it should seem not: for that which is used to prove every thing else, cannot itself be proved: a chain of proofs must have their commencement somewhere. To give such proof is as impossible as it is needless.¹⁷

There can be no doubt that both Bentham and Mill accepted this axiom wholeheartedly and consistently throughout their entire works. I have deliberately distinguished between the principle of utility and the proposition that it is *only* pleasure that has value for human beings, because the latter addendum was never doubted by Bentham whereas Mill labored much harder to convince himself of its validity. The proposition that underlies the uniqueness of hedonic value, according to my interpretation of classical utilitarianism, will therefore be stated separately as axiom three.

The next two axioms I would classify as ontological, being assertions about the nature of reality. Although I find them crucial to a proper understanding of the ethical system that Bentham and Mill conceived, they have, I believe, been largely overlooked by subsequent commentators. The first of these, my second axiom of four, sums up the basic contention of Section Two: **Axiom 2.** Happiness for a human being is irreducibly multi-dimensional, with a distinct dimension for each type of pleasure and for each type of pain. The crucial qualifier here is ‘irreducibly’, for no one can doubt that pleasure and pain take many different forms. The issue is whether there exists some *generic* pleasure that can serve as a common denominator for all its types

¹⁷ *IPML (CW)*, p.13.

and, with negative value, for every species of pain. My argument is that both Bentham and Mill consistently denied commensurability among separate forms of pleasure and pain. The textual evidence for my position is presented at length in ‘Multi-Dimensional Utility’, from which I shall quote just two examples.¹⁸ From Bentham: ‘The elements of happiness are *pleasures* and *exemptions from pains*: individual pleasures, and exemptions from individual pains’.¹⁹ And from Mill: ‘The principle of utility does not mean that any given pleasure, as music, for instance, or any given exemption from pain, as for example health, are to be looked upon as a means to a collective something termed happiness, and to be desired on that account’.²⁰ Unfortunately, Bentham and Mill did *not* have at their disposal the concept of a pleasure/pain vector, nor of utility as an index number, and were thus unable to employ what are in fact the only rigorous representations of the irreducible multi-dimensionality of happiness that I am certain they took for granted.

The second of my ontological axioms is the addendum to the principle of utility that excludes anything other than experienced pleasure or pain as an indicator of value: **Axiom 3.** Any attribution of intrinsic worth to any form of human behavior is a human artifact. I use ‘intrinsic worth’ here in the same sense as did Moore, meaning an attribute, or named aspect, of an act or state of mind that is recognizably ‘good’ independent of the response of any agent, and which therefore ought to exist for its own sake:

Every one does in fact understand the question “Is this good?” When he thinks of it, his state of mind is different from what it would be, were he asked “Is this pleasant, or desired, or approved?” It has a distinct meaning for him, even though he may not recognize in what respect it is distinct. Whenever he thinks of “intrinsic value”, or “intrinsic worth”, or says that a thing “ought to exist”, he has before his mind the unique object - the unique property of things - which I mean by “good”.²¹

Had he encountered this passage, Bentham might well have accepted it as a valid description, not for ‘every one’, but at any rate for many persons’ states of mind

¹⁸ Warke, ‘Multi-Dimensional Utility and the Index Number Problem’.

¹⁹ *Legislator of the World (CW)*, p. 251.

²⁰ Mill, ‘Utilitarianism’, p. 235.

when considering the meaning of good. He would also, I am convinced, have gone on to assert that this phenomenon of human thought, so accurately described by Moore, is the single greatest *impediment* to the implementation of a rational (i.e., utilitarian) system of ethics:

Among principles adverse to that of utility, that which at this day seems to have most influence in matters of government, is what may be called the principle of sympathy and antipathy. By the principle of sympathy and antipathy, I mean that principle which approves or disapproves of certain actions, not on account of their tending to augment the happiness, nor yet on account of their tending to diminish the happiness of the party whose interest is in question, but merely because a man finds himself disposed to approve or disapprove of them: holding up that approbation or disapprobation as a sufficient reason for itself, and disclaiming the necessity of looking out for any extrinsic ground.²²

The philosophers who, in my view, have persistently misinterpreted classical utilitarianism have generally accepted without question Moore's grounds for debate, namely whether or not it is valid to regard pleasure and only pleasure as having intrinsic value. By my interpretation, when Bentham said, 'Now, pleasure is in itself a good: nay, even[,] setting aside immunity from pain, the only good', he was affirming what I have called his Euclidean axiom, using 'good' in the sense that Moore rejected - a general quality of any experience perceived as pleasant - and *not* in the Moorean sense of intrinsic value.²³ If, as I believe, Bentham accepted my axiom two, then he could only have been referring to a quality that all forms of pleasure share, and not to a quality of a generic pleasure that does not exist. Whenever Bentham *specified* potential sources of pleasure - pushpin or poetry, benevolence or malevolence - he consistently asserted that their value or disvalue was acquired solely by the responses they excited in sentient beings. This distinction between Bentham's hedonic value and Moore's intrinsic value was, the reader will recall, a central point of my illustration in Section Two. In that formulation, a dose of beauty or amity or impiety had no

²¹ Moore, *Principia Ethica*, pp. 16-17.

²² *IPML (CW)*, pp. 21-25.

²³ *Ibid.*, p. 100.

intrinsic value, positive or negative; a utilitarian's approval or disapproval of an act can *not* 'disclaim the necessity of looking out for an extrinsic ground', for the value of a dose remains latent unless and until it is accorded a non-zero weight in the response vector of at least one sentient being within its sphere of influence.

Compared to Bentham's certainty and consistency on this point, the extent to which Mill accepted my third axiom is more problematic, and worthy of more extensive comment than I can give it here. In 'Multi-Dimensional Utility and the Index Number Problem', I argue that he managed, just barely, to avoid its abrogation in *Utilitarianism*, and I suggest that this flirtation with intrinsic value was motivated mainly by the polemic purposes of his 1861 essay. In the 1830s he almost certainly would have denied the axiom. Furthermore, Mill never showed the scorn that Bentham did for the classical Greek philosophers, in their search for intrinsic value among types of behavior and character traits. But by my reading of *Utilitarianism*, Mill had at the time of its writing returned, perhaps reluctantly, to the Benthamite fold with regard to axiom three. Thus, in Chapter II's exaltation of intellectual pleasures, Mill stretched to the limit, but resisted straying beyond, assertions of *empirical* evidence to support their supposed superiority vis-à-vis animal pleasures.

The last of my axioms I shall term operational, as it is required to transform the principle of utility into a determinate standard of right and wrong. I refer, of course, to egalitarian interpersonal weighting, and can think of no better phraseology than Bentham's own: **Axiom 4.** Everybody to tell for one, and nobody for more than one.

In one sense, egalitarian interpersonal weighting dominated Bentham's thought, for it eventually turned him into a radical advocate of representative democracy as a prior condition for his utilitarian agenda. Nonetheless, he was sometimes slipshod in his allegiance to egalitarianism, especially in his earlier work. For example, some passages of *An Introduction to the Principles of Morals and Legislation* sacrifice equality under the law to efficacy of punishment, according to which the severity of statutory penalty pains should be related to the strength of temptations to mischief. Thus, although he showed considerable discomfort about the implied violation of equality, Bentham nonetheless proposed as a criterion for punishment of theft that it be inversely related to the wealth of the miscreant. I would attribute Bentham's inconsistency on egalitarianism to his lack of a rigorous framework for evaluating aggregate happiness, for he did not always recognize that the allocation of interpersonal weights is a step that is distinct from (and a constraint upon) the over-

arching proposition that maximum aggregate happiness is the right and proper end of human conduct.

If Bentham was sometimes careless, Mill was often cavalier with interpersonal weighting. This accords in part with Mill's greater interest in personal choice, compared to Bentham's concentration on public choice. Even for a topic that was clearly in the public sphere, such as justice, Mill was more concerned with its implications for individual autonomy than with finding a Benthamite balance between the pain imposed on perpetrators of mischief and the pain suffered by their victims. In addition, however, it seems to me that Mill was more prone than Bentham to making ad hoc adjustments on interpersonal weights in order to secure a result in agreement with his own tastes and preferences. This sort of departure from egalitarianism is especially blatant within one well-known Millian exercise in evaluation, to which I have already alluded:

What is there to decide whether a particular pleasure is worth purchasing at the cost of a particular pain, except the feelings and judgment of the experienced? When, therefore, those feelings and judgment declare the pleasures derived from the higher faculties to be preferable *in kind*, apart from the question of intensity, to those of which the animal nature, disjoined from the higher faculties, is susceptible, they are entitled on this subject to the same regard.²⁴

As W. S. Jevons remarked: 'The verdict which Mill takes in favour of his high-quality pleasures is entirely that of a packed jury. It is on a par with the verdict which would be given by vegetarians in favour of a vegetable diet'.²⁵

To summarize this section, I contend that classical utilitarianism is defined by the first three of the axioms I have listed, and that Bentham, and Mill of *Utilitarianism*, remained faithful to them. Although Bentham and Mill each departed at times from my fourth axiom, I believe that Bentham, at least, would have acknowledged a precedence of egalitarian interpersonal weighting had the inconsistency with it of some of his proposals been brought to his attention. Bentham

²⁴ Mill, 'Utilitarianism', p. 213.

²⁵ W.S. Jevons, 'John Stuart Mill's Philosophy Tested, IV – Utilitarianism', *Contemporary Review*, xxxvi (1879), p. 532.

had a strong mathematical intuition and would, I believe, have been quick to recognize, firstly that ethical determinations within a formal system must *use* interpersonal weights, and secondly *consistent* ethical determinations call for a pre-specified set independent of any particular application. Once those facts were recognized, I have little doubt that Bentham would have specified equality.

I will conclude this section with a personal view on the definition of utilitarianism more generally. By my lights, it is the first and third of my axioms that distinguish utilitarian from non-utilitarian ethics, and I believe that this tracks with Bentham's threefold classification of ethical systems: Bentham's principles of utility and of asceticism correspond respectively to acceptance of axiom 1 and acceptance of the converse of axiom 1; Bentham's principle of sympathy and antipathy corresponds to the negation of axiom 3. According to this definition, the utilitarian umbrella casts a copious penumbra. Included within it is most of welfare economics, despite its rejection of axiom 2. A system that replaces axiom 4 with a specified alternative interpersonal weighting is not excluded and, so far as I have read him at present, that brings John Rawls into the camp. Based on my reading of Warnock's *Ethics since 1900*, I see no reason to reject Ayer and Stevenson. Yet despite the breadth of my definition, I would most firmly classify Moore's *Principia Ethica* as non-utilitarian.

4. Mill on Virtue

For the final substantive section of this paper, I shall comment very briefly upon certain aspects of Mill's treatment of virtue in *Utilitarianism*. The four axioms that I presented in Section Three are a basis for ethical hedonism, but do not address the issue within psychological hedonism that I raised initially in Section Two: Are there ever circumstances in which an agent would undertake voluntarily an act from which he or she did not expect a net pleasure advantage for himself or herself? Bentham's answer was an emphatic 'no', a position to which Mill took strong exception in his 1830s analysis of conscience. Mill's discussion of virtue in *Utilitarianism* returns to the same question, and reverts partially but not entirely to a Benthamite stance, as I hope to demonstrate below. In any case, Mill's treatment of the subject in 1861 is much more sophisticated, and psychologically richer, than anything Bentham had offered. The reason my comments can remain brief is that Berger covers the same ground extremely well, and with whom, on this topic at least, I am in substantial agreement.

For both Bentham and Mill, virtue was a quality of acts whose consequences increase the happiness of sentient beings other than the actor. According to Bentham's classification of motive, virtue would be a *by-product* - a positive externality - of actions that are induced by certain types of potential pleasure from his supposedly-exhaustive catalog, such as amity or benevolence. Although he acknowledged, and sometimes stressed, that an individual's sensibility to pleasures with positive externalities could be enhanced by education, by good example, and by sociability in general, Bentham for the most part concerned himself with the design and promotion of *sanctions* that would induce virtuous acts or deter non-virtuous acts through the threat of potential pain. In addition to political sanctions, imposed via the judicial system, the force that Mill called 'conscience' was subsumed by Bentham within the moral sanction: agents undertake virtuous acts despite a net pain balance to themselves if and only if they thereby avoid a greater pain from loss of reputation.

Mill's most substantive amendment to Bentham's schema was the notion that an agent might undertake an act just because he or she perceived it to be the right thing to do, or might avoid an act just because he or she perceived it to be the wrong thing to do, irrespective of the net pleasure balance expected to be experienced from doing the act. In the 1830s, Mill named this overriding motive 'conscience'. In *Utilitarianism*, he incorporated conscience within psychological hedonism by introducing potential pain or pleasure from the *idea* of performing certain acts, as a supplement to expected pains and pleasures from the performance itself. In particular, Mill argued that the idea of virtue in one's choices can become a source of pleasure:

Whatever may be the opinion of utilitarian moralists as to the original conditions by which virtue is made virtue; however they may believe (as they do) that actions and dispositions are only virtuous because they promote another end than virtue; yet this being granted, and it having been decided, from considerations of this description, what *is* virtuous, they not only place virtue at the very head of the things which are good as means to the ultimate end, but they also recognize as a psychological fact the possibility of its being, to the individual, a good in itself, without looking to any end beyond it.²⁶

In terms of my formal schema in Section Two, Bentham's position was, firstly, that education, example and sociability could increase the relative intensities that an

²⁶ Mill, 'Utilitarianism', p. 235.

agent allocates to virtue-inclusive pleasures such as amity and benevolence (or could reduce the relative intensity allocated to the pleasure of malevolence), and secondly, that utilitarian legislators and moralists could attach new types of pain doses to acts that would otherwise confer a net pleasure balance on the perpetrator, but which impose a larger net pain balance on other agents. What Mill added was the possibility that virtue could itself become a pleasure dose to which agents allocate a positive response. If this occurs, such persons would have a motive (as opposed to an obligation) to examine the consequences of acts for other people in order to know what are the virtue doses the acts entail for themselves. In other words, virtuous persons are agents who incorporate the general happiness as a part of their own happiness. In my example, if Ben is virtuous and Geri is not, he would note her net pleasure gain from Act M and would himself take pleasure at the idea of catering for it, whereas her evaluation of the two alternatives would remain unchanged. If the intensity weight that he allocates to the pleasure of virtue is sufficiently high, then Act M would become the mutually-agreed choice.

In contrast to the situation in Section Two where Ben's conscience had to override his psychological hedonism in order to agree to Act M, Ben's virtue now brings his psychological hedonism into accord with Geri's. An observing Benthamite, once persuaded (presumably by empirical evidence) that Bentham's catalog of potential pleasures should be amended to include virtue, would this time designate their joint choice of Act M as right. Of course, if neither Geri nor Ben allocate a positive weight to virtue, Act A would be right as before. It should be noted, however, that the *available* pleasure from either act is reduced if virtue makes no contribution. A virtuous person will generally be a happier person than a non-virtuous person so long as it is a pleasure gain from virtue that tips the balance compared to self-regarding choices (as opposed to aversion from potential pain that the virtuous person attaches to self-indulgent acts, thus bringing their net pleasure balance below the virtuous acts).

This last observation illustrates the point, mentioned previously, that egalitarian interpersonal weights are a *constraint* on the greatest happiness principle. If the utilitarian end were unconstrained, persons who are more susceptible to happiness would be weighted more heavily in the hedonic calculus than persons who are less susceptible to happiness. Over and above any enhanced susceptibility to happiness, virtuous persons would receive a further supplement to their weighting within an

unconstrained greatest happiness principle, due to the positive externalities of their choices. Of course, the implementation of such weighting supplements in public choice would require that legislators or judges designate who is virtuous and who is not, and this is a can of worms that egalitarian weighting avoids. It is because a utilitarian case can be made on either side of the issue, however, that I do not regard my fourth axiom as *necessary* in order to define an ethics as utilitarian. Put another way, I would dissent from Hare's view that egalitarian interpersonal weighting is somehow inherent in the logic of moral reasoning.

Thus far in my discussion, Mill's treatment of virtue in Utilitarianism has enlarged the scope for harmony of interests between psychological and ethical hedonism. So long as virtue is accepted as a potential pleasure (and/or lack of virtue as a potential pain), it remains true that an agent would not undertake voluntarily an act unless he or she expected a net pleasure advantage for himself or herself. However, Mill went on to present a counterexample to psychological hedonism as the determining factor for individual choice. It is possible, said Mill, for virtue (or any other purpose) to become a *habit*, so that it still induces actions even when it no longer confers pleasure. In Mill's terms, so long as desire determines choice, psychological hedonism holds sway by definition, but for a person 'whose purposes are fixed [...] instead of willing the thing because we desire it, we often desire it only because we will it.'²⁷ Mill himself, however, emphasized two points that are very relevant to this paper. First, he insisted: 'It is not the less true that will, in the beginning, is entirely produced by desire; including in that term the repelling influence of pain as well as the attractive one of pleasure'.²⁸ Thus, as Berger phrased it:

[Mill] maintained that pleasure and pain are *causally* linked to all voluntary human acts. He did not hold, however, that all acts are done with the thought of gaining a pleasure by means of them.²⁹

Mill's second point about a virtuous will supplies a fitting conclusion for this paper, for it demonstrates once again his acceptance in *Utilitarianism* that no specific

²⁷ *Ibid.*, p. 238.

²⁸ *Ibid.*, p. 238.

²⁹ Berger, *Happiness, Justice and Freedom*, p. 239.

type of pleasure, not even virtue, has intrinsic value in Moore's sense. Habitual virtue, where the perpetrator of virtuous acts attaches zero weight to virtue as pleasure, would be without value *except* for the positive externalities that make the acts virtuous - experienced pleasures or pain aversions for other persons. Because habitual virtue has less hedonic value than virtue-as-pleasure, the only reason why a utilitarian should welcome its existence is its 'constancy of action', the psychological reality that habitual virtue will produce a larger *number* of acts with positive externalities than will virtue-as-pleasure:

That which is the result of habit affords no presumption of being intrinsically good; and there would be no reason for wishing that the purpose of virtue should become independent of pleasure and pain, were it not that the influence of the pleasurable and painful associations which prompt to virtue is not sufficiently to be depended on for unerring constancy of action until it has acquired the support of habit. [...] In other words, this state of the will is a means to good, not intrinsically a good; and does not contradict the doctrine that nothing is good to human beings but in so far as it is either itself pleasurable, or a means of attaining pleasure or averting pain.

Conclusion

There are two main themes of this paper, as I have conceived it. The first is my contention that utilitarian ethics, classical or contemporary, is defined by two axioms: (i) it is better for a sentient being to experience pleasure than pain; and (ii) any attribution of intrinsic worth to any form of human behavior is a human artifact. The second theme is my contention that the classical utilitarians can not be understood properly unless it is recognized that their utility concept was irreducibly multi-dimensional, and that it is therefore useful to reconstruct their system with individual happiness as an index number, consisting of pleasure/pain doses weighted by the individual's corresponding set of relative intensity responses. The degree to which the paper supports its themes is, of course, for the reader to judge.

Copyright © 1999, Tom Warke. This file may be copied on the condition that the entire contents, including the header and this copyright notice, remain intact.