

FROM UTILITARIANISM TO KANTISM: BENTHAM'S PROOF OF UTILITARIANISM, MILL AND KANT¹

JOSÉ DE SOUSA E BRITO*
Universidade Nova de Lisboa

ABSTRACT

Bentham has given different formulations of the principle of utility and also tried in different ways to prove it. In his later writings, some of them unpublished, he makes a distinction between two senses of the principle of utility: the exegetic or expository and the deontological or censorial. They are indeed two different principles leading to different results, one egoistic, altruistic the other. The legislator ought to make them artificially coincide by using the proper sanctions. Mill's "proof" of the principle is a misleading attempt to derive the deontological from the expository sense. Bentham's attempts to prove the deontological principle of utilitarianism bring Bentham closer to Kant than usually admitted.

Keywords: Immanuel Kant, utilitarianism, John Stuart Mill, Jeremy Bentham, deontologism.

RESUMEN

Bentham ha proporcionado diferentes formulaciones del principio de utilidad y también ha ensayado diversas formas de probarlo. En sus últimos escritos, algunos de ellos inéditos, formula una distinción entre dos sentidos del principio de utilidad: la exegética o expositoria y la deontológica o censoria. Ciertamente se trata de dos principios que conducen a diferentes resultados, uno egoísta y el otro, altruista. El legislador debe hacerlos coincidir de un modo artificial usando las sanciones adecuadas. La "prueba" del principio de Mill es un intento confundente de derivar el sentido deontológico del sentido expositivo. El intento de Bentham de probar el principio deontológico del utilitarismo sitúa a Bentham más cerca de Kant de lo que es normalmente aceptado.

Palabras clave: Immanuel Kant, utilitarismo, John Stuart Mill, Jeremy Bentham, deontologismo.

¹ I thank Fred Rosen for his comments on an earlier version of this paper.

* Date of acceptance: 26/07/2006.

1. IS THE PRINCIPLE OF HAPPINESS DERIVABLE FROM REASON? KANT V. BENTHAM

Kant opposed his own moral theory to the theory of happiness and so to utilitarianism –before the word existed– in the strongest terms. “If *eudemony* (the principle of happiness) –he says– is put forward as a principle instead of *eleutheronomy* (the principle of freedom of interior legislation) , the consequence of it is the *euthanasia* (the gentle death) of the whole morality”². Kant refuses to admit that happiness be the ideal of reason³ and that any theory of happiness can be derived from *a priori* reasoning. His argument for the later thesis runs as follows: “only experience can teach us, what brings us joy. The natural impulses to food, to sex, to rest, to move, and (by developing our natural capacities) the impulse to honour, to the extension of our knowledge, etc. can only make known to each one in his peculiar way, where he shall find such joy, and the same can also teach him the means by which he shall pursuit it. Any apparent reasoning *a priori* is here basically nothing more than experience elevated through induction to universality”⁴.

It is therefore thought-provoking when Richard Hare, who acknowledges that he owes so many of his own ideas to Kant⁵, pretends to reconcile Kant with utilitarianism: “the formal, logical properties of the moral words –so Hare–, the understanding of which we owe above all to Kant, yield a system of moral reasoning whose conclusions have a content identical with that of a certain kind of utilitarianism”⁶. By saying this, Hare joins no less than John Stuart Mill, who departs from the opposite utilitarian side. According to Mill, the general principles of Teleology, or the Doctrine of Ends, which are for him those of utilitarianism, “borrowing the language of the German metaphysicians –here Mill is referring to Kant–, may also be termed, not improperly, the principles of Practical Reasoning”⁷. But Mill does not say how utilitarianism can eventually turn out to be properly Kantian. When he returns later to the point, he simply states that Kant’s categorical imperative can only make sense in line with some kind of consequentialism⁸. Richard Hare tries equally

2 *Metaphysik der Sitten, Tugendlehre*, 1797, A IX (Akademie-Ausgabe 378).

3 *Grundlegung der Methaphysik der Sitten*, 1785, BA 47 (Akademie-Ausgabe 418).

4 *Metaphysik der Sitten, Rechtslehre*, 1797, AB 9 (Akademie- Ausgabe 215-216).

5 *Sorting out Ethics*, Oxford, Clarendon Press, 1997, p. VI; see also p.99: “I think, I am still a follower of Kant”.

6 *Moral Thinking*, Oxford, Clarendon Press, 1981, p.4. Hare developed later the same thesis in his paper “Could Kant have been a Utilitarian?” (1993) in *Sorting out Ethics*, p. 147 ff. I am asking in this paper the opposite question: - Could, or rather should Bentham be considered a Kantian *avant la lettre*?

7 *A System of Logic*, book VI, ch. Xii, § 6 (1851) (ed. J.M. Robson in the *Collected Works*, Toronto, University of Toronto Press, 1974, pp.949-950).

8 *Utilitarianism* (1861), in *Essays on Ethics, Religion and Society*, ed. J. M. Robson in the *Collected Works*, X, 1969, pp. 207, 249.

to interpret Kant as being compatible with utilitarianism on the same and on other points⁹.

However, the true road from utilitarianism to Kantism was laid down by Bentham even before Kant wrote the *Groundwork of the Metaphysics of Morals*. To substantiate this claim, it is necessary to expose and to discuss Bentham's proof of utilitarianism, to interpret in his light the most general principles of utilitarianism and to compare with Kant's categorical imperative. A contribution to all of that will be attempted in what follows.

To compare the conclusions arrived at by Kant with those of utilitarianism, as Hare earlier proposed, does not seem advisable. First because there is not much hope of agreement. As Hare later recognized¹⁰, Kant's puritanical views on capital punishment, suicide, lying, civil disobedience, for example, would not be shared by Bentham or Mill or almost any modern moralist. Secondly and mainly, because there should be no difference in the conclusions if there is indeed a common departure of the reasoning.

2. FROM PRINCIPLE TO PROOF AND VICE VERSA

The close relation between the content of a theory and its proof is suspiciously affected in the case of utilitarianism by the dogmatic rigidity of the formula "the greatest happiness of the greatest number" that characterized the theory since it became a banner for political reform. Bentham himself did not resist always the dogmatic temptation. This is revealed by a significant story. We know that Bentham was very much afraid of nightmares¹¹. So we are slightly astonished to see that such a lover of daily truth indulged once in telling us with complaisance a dream he had in 1780 aged 32: "I dreamt the other night that I was a founder of a sect; of course a personage of great sanctity and importance It was the sect of utilitarians"¹². Commenting on this Bentham admitted that this was mad but he reassured us that his "madness has not yet gone beyond a dream". The dream became almost true much later but it remained philosophically mad. By being invested with religious glory the principle of utility turned into a dogma. A dogma should neither be criticised nor changed. It does not depend on proof but on authority. For that reason it probably proves itself useful for pushing a reformist sect for action. So

9 See Richard Hare, "Could Kant have been a Utilitarian?", specially pp. 163 ff.

10 "Could Kant have been a Utilitarian?", p. 148.

11 See Bowring in John Bowring (ed.), *The Works of Jeremy Bentham*, 1838-43 (reprint New York, Russell, 1962), XI, p. 82.

12 Bentham MSS in the Library of the University College, London (U.C.) CLXIX,79:in David Baumgardt, *Bentham and the Ethics of today*, Princeton, Princeton University Press, 1952, App. I and Parekh (ed.), *Bentham's Political Thought*, London, Croom Helm, 1973, pp. 13 and 325.

we shall find the founder of the sect in the short version of the "Article on Utilitarianism", intended for the *Westminster Review*, sticking to the old formulation of the principle of utility - "greatest happiness of the greatest number" - just after having demonstrated in the long version how some years before "reason altogether incontestable, was found for discarding this appendage"¹³, namely, "of the greatest number".

This brings us to the relation between principle and proof. The proof may come after the principle in the way of discovery. It certainly comes first in the way of demonstration. Not only the formulation of the principle depends on the proof but also the whole system of derived principles and applications, i.e. utilitarianism as a whole and not only its principle depends on its proof. This does not exclude a dependence of the principle also on its derivations. The reason for discarding the appendage "of the greatest, number" was the following: the principle of acting for the greatest happiness of the greatest number was a derivation of the simple greatest happiness principle in cases of conflict between the happiness of "the few" and the happiness of "the many"¹⁴ and was a derivation with unacceptable applications. I quote: "Be the community in question what it may; divide it into two unequal parts, call one of them the majority, the other the minority, lay out of the account the feelings of the minority, include in the account no feeling but those of the majority, the result you will find is that to the aggregate stock of the happiness of the community, loss not profit is the result of the operation... note now the practical application that would be made of it in the British Isles. In Great Britain, take the whole body of the Roman Catholics, make slaves of them and divide them in any proportion, them and their progeny among the whole body of the Protestants. In Ireland, take the whole body of the Protestants and divide them alike among the whole body of the Roman Catholics"¹⁵. Why are they unacceptable? That depends on proof. If Bentham had been a kind of intuitionist, the immediate intuition of the wrongfulness of a specific application would correct the import given to a principle on the basis of another intuition. In spite of some rhetorical appeal to intuition, Bentham does not abide by it. So he speaks of an "incontestable reason" for discarding the appendage and indeed in the "Pannomial Fragments" he gives a new formulation to the appendage in the form of "the next specific principle" (i.e. next to the greatest happiness principle): "the happiness numeration principle": "In case of collision and contest: *happiness of each party being equal* prefer the happiness of the greater to that of the lesser number". This formulation avoids the former applications in which the happiness you take

13 *Deontology*, ed A. Goldworth, Oxford, Clarendon Press, 1983 (*The Collected Works of Jeremy Bentham*), p. 309.

14 "Pannomial Fragments", Bowring, III, p. 211.

15 *Deontology*, p. 310.

away has more value than the happiness you give. So what one should do is to reconstruct utilitarianism, departing from its possible proofs and to check the reconstruction against the textual evidence. Because of the sectarian infection of the late Bentham, one should not rely too much on chronology but concentrate on systematic interpretation of the main texts. As one cannot do systematic interpretation without some rational reconstruction, both go hand in hand. John Stuart Mill thought, I suggest, much in the same lines, when he wrote: "a doctrine is not judged at all until it is judged in its best form"¹⁶.

This is a vast program. In my paper I shall try to show how to begin the procedure. I shall start with what I take to be the best formulation of the greatest happiness principle, revealing some of its systematic links in a view to demonstrate its superiority to alternative formulations. After setting thus the theme of proof, I shall try to interpret systematically what Bentham says about the proof. Instead of going on to a full rational reconstruction of the proof first and of utilitarianism after, I shall only show some new perspectives, which are opened by the confrontation between the new interpretation of the first principles of utilitarianism that will follow from the rational reconstruction of its proof and both the traditional version of utilitarianism and the corresponding Kantian criticism.

3. BENTHAM'S TWO PRINCIPLES OF UTILITY: EXPOSITORY AND DEONTOLOGICAL

In the spring of 1814 Bentham started writing about the Chrestomathic School. This may well have been the practical sting. Anyhow, during three years between mid 1814 and mid 1817, Bentham wrote mostly about philosophical questions: on ethics, logic, language, grammar and ontology. In no other period of his life did he concentrate so much on those matters. The published outcome consisted of two small, extremely dense books: *A Table of the Springs of Action* (1815-17) and Part II of *Chrestomathia* (1816) consisting of an Appendix, the "Essay on Nomenclature and Classification". They constitute with *An Introduction to the Principles of Morals and Legislation* (1780-89) Bentham's masterpieces of Philosophy. Actually the publication of the three passed almost unnoticed. Bentham was in his mid sixties, his style was somewhat pedantic, compared with a quarter of century before, but the standards of writing were even higher. On 13 July 1815 Bentham assembled the marginals of an intended extension of the already printed main text of *A Table of the Springs of Action*. There we read: "Principles of utility two, or if but one, it is understood in two senses - viz. The censorial and the expository or exhibitivive (Censorial, what. Expository, what)"¹⁷. We don't have

16 "Sidgwick's Discourse" (1835), *Collected Works*, X, p.52.

17 U.C. CLVIII, 105 (=Deontology, p.9: Goldworth reads "expositive or exhibitivive").

the corresponding text, but we have another one, written on 10 April 1816, perhaps as a substitute to the former text, eventually destroyed by Bentham¹⁸. It is entitled: "Principle of Utility - its two senses", and it has unhappily not been published with *A Table* in the *Collected Works*. It begins so: "The noun or phrase principle of utility is in use to be employed in two different senses: the Exegetic or Expository, and the Deontological or Censorial.

These two senses are so far allied that for the designation of both of them one and the same locution is needed to be employed: but they are so far different and the difference is to such a degree essential that to him who in a subject of such all-comprehensive importance would wish to know what he reads or says or thinks, it is highly important that of this difference a clear conception should in every occasion be entertained".

Let me now transcribe directly the definitions. In the expository sense, the principle would have this formulation: "of every human being the conduct is on every occasion determined by the conception which at that moment he has of his individual interest"¹⁹. Interest is understood in its "largest sense", embracing all kinds of interests - corresponding to all kinds of pleasures, desires and motives-, not only those of the "self-regarding" class, but in Bentham's words, also the "social" and the "dissocial" ones. In the deontological sense the principle has a very different import. I quote it from the same manuscript that I was using before - apparently the only extant one where the distinction is fully developed: "it is desirable - fit, right, proper, desirable - every one of these words may be employed, that on every occasion the course taken by every man's conduct should be that which will be in the highest degree conducive to the welfare of the greatest number of those sensitive beings on which welfare it exercises any influence"²⁰.

Comparing this formulation with that of *An Introduction to the Principles of Morals and Legislation*, there are some main differences:

1. It refers to the conduct that should be taken instead of the approval or disapproval you put on an action. A practical principle is in first place a guide to conduct and only consequently a guide to moral evaluation of it. "Conduct" is more general than "action", it applies also to omissions. The talk about tendency in *An Introduction* led some linguistic philosophers to see in Bentham a rule-utilitarian *avant la lettre*, because only types of action have, properly speaking, a tendency. That you can only foresee

18 U.C. XVIII, 172-175.

19 U.C. XVIII, 173.

20 *Ibid.* See Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation*, ed. J. H. Burns, H. L. Hart, London, Athlone Press (*Collected Works*), pp. 11-12: "By the principle of utility is meant that principle which approves or disapproves of every action whatsoever, according to the tendency which it appears to have to augment or diminish the happiness of the party whose interest is in question".

the outcome of action on the basis of causal laws must not conceal that the principle guides individual conduct and not types of conduct.

2. "Welfare" is taken as synonym of "Well-being", which is the word Bentham ultimately preferred to oppose to happiness. In *Deontology* he wrote: "Instead of well-being the word "happiness" will not be equally suitable to the purpose. It means not only to lay pain in all its shapes altogether out of the account, but to give it to be understood that whatever have been the pleasures that have been experienced, it is in a high and as it were superlative degree that they have been experienced"²¹.

3. "Sensitive beings" renders more accurately Bentham's, thinking even at the time of *An Introduction* where he first fought for animal rights: "The day may come where the rest of the animal creation may acquire those rights which never could have been withholden from them but by the hand of tyranny... the question is not, Can they *reason*? nor, Can they *talk*? but, Can they *suffer*?"²².

4. The formulation of *An Introduction* is superior in omitting the appendage "of the greatest number" for the reason exposed earlier in this paper.

5. The reference to the influence of conduct on welfare or well-being makes explicit Bentham's consequentialism²³, whereas the talk about interests being in question in *An Introduction* could be misunderstood as the questioning of interests, varying the moral criteria according to the number of persons considered and not to those that are affected.

6. The common designation "principle of utility" is just as bad as "greatest happiness principle", the first does not point out the specific utility endeavoured by the deontological principle, the latter is open to the objections related to well-being.

7. The formulation of *An Introduction* covers only the deontological principle.

In this way the resulting interpretative, not yet entirely rational, reconstruction of the deontological principle would be something like this: on every occasion the course taken by every man's conduct should be that which will in the highest degree be conducive to the well-being of those sensitive beings on which well-being it exercises any influence.

4. THE PROOF OF THE EXPOSITORY PRINCIPLE

We have now two different principles to prove and more than that, two different kinds of proof.

21 P. 130.

22 P. 282-3.

23 Cf. *An Introduction*, p.74.

Bentham deals with the proof of the expository principle when trying to prove one of its applications, the second axiom of government, according to the "Constitutional Code Rationale" of 1822²⁴: "The actual end of government is in every political community the greatest happiness of those, whether one or many, by whom the powers of government are exercised. In general terms the proof of this position may be referred to particular experience, as brought to view by the history of all nations." This is, I think, empirically wrong, but conceptually perfect: it could only be, if ever, empirically proved. But Bentham moves further to the proof of the more general expository principle which was at the time denominated by Bentham's self-preference principle²⁵. He says: "For further proof reference may be made to the general indeed the all-comprehensive, principle of human nature ... In the general tenor of human life, in every human breast, self-regarding interest is predominant over all other interests put together. More shortly thus, self-regard is predominant or thus, self-preference has place every where"²⁶. One would expect him to proceed: the proof of this position may be referred to particular experience, as brought to view by the history of mankind. At this point, however, Bentham goes astray. He tells us astonishingly: "of this stamp are the axioms laid down by Euclid" and he goes on to a kind of counterfactual *reductio ad absurdum*: "reference may be made to the *existence* of the species as being of itself a proof - and that a conclusive one. For, after exception made of the case of children not arrived at the age at which they are capable of going alone, or adults reduced by infirmity to a helpless state, take any two individuals, A and B, and suppose the whole care of the happiness of A to be confined to the breast of B, A himself not having any part in it, and the whole care of the happiness of B confined to the breast of A, B himself not having any part in it - and this to be the case. Throughout, it will soon appear that in this state of things the species could not continue in existence, and that a few months, not to say weeks or days, would suffice for the annihilation of it"²⁷.

It may be asked if the newly called principle of self-preference is identical with the expository principle of utility. The conception which someone has of his individual interest can be predominantly determined

24 In *First Principles Preparatory to Constitutional Code*, ed. P. Schofield, Oxford, Clarendon Press, 1989 ("Collected Works"), pp. 232 ff. This part of the text was already published by Doane in Book I of the *Constitutional Code*: Bowring, IX, pp. 5-6. See Bowring, X, p. 80.

25 Cf. Bowring, IX, p.5: "By the principle of self-preference, understand that propensity in human nature, by which, on the occasion of every act he exercises, every human being is lead to pursue that line off conduct which, according to hid view of the case, taken by him at the moment, will be in the highest degree contributory to his own greatest happiness, whatsoever be the effect of it, in relation to the happiness of other similar beings, and, or all of them taken together". Cf. the other formulation in Bowring, X, p. 80: "man, from the very constitution of its nature, prefers his own happiness to that of other sensitive beings put together".

26 I prefer the reading of Doane.

27 *First Principles Preparatory*, p. 233.

by a social interest which is by definition extra-regarding. But Bentham says about the interest corresponding to or produced by the affection of sympathy or benevolence: "This it is true is an extra-regarding interest, but it is not the less a self-regarding one"²⁸. Thus we have a sense of "self-regarding" which excludes "social" and another which embraces it. The last one should be meant by the principle of self-preference. This is confirmed by a text of 1822 where we read that by this principle "neither the tenderest sympathy nor anything that commonly goes by the name of disinterestedness, improper and deceptive as the appellation is, is denied". Sympathy, Bentham admits it of any pain or pleasure, "may have magnitude enough in the mind to eclipse all other pains, as well as all other pleasures"²⁹. The best example is Bentham himself for whom the establishment of the principle of utility would in the highest degree be contributory to his own greatest happiness³⁰. If so, the principle cannot be properly expressed as stating the predominance of self-regarding interest over all other interests of *the same person*. Does the argument from the existence of mankind prove anything? If the argument holds, it would simply follow that the principle contrary to the principle of self-preference, i.e., the principle of the preference for others would lead to a result that cannot be wished, the destruction of mankind. This is a practical absurdity, but it does not entail a contradiction, it is not a *reductio ad absurdum* of the geometrical kind.

How can then the principle of self-preference or principle of utility in expository sense be based on experience? If at least one person sometime acts against his own view of his greatest happiness, the principle is empirically false. And the cases of weakness of the will are instances of just that. It may still remain as practical rule that is followed in most cases. This interpretation of the principle may be conciliated with its admission as an empirical probability. Bentham speaks about the last when he says: "note that if in the situation of a ruler the truth of this position (i.e., self-preference) held good in no more than a bare majority of the whole number of instances, it would suffice in the character of a ground for practice: in the character of a consideration by which the location of the several portions in the aggregate mass of political power should be determined"³¹. There is, however, a deeper sense in which the expository principle can be said to be based on experience. It reveals the structure of experience and indeed of human action as intentional and for that rational action. Even in case of weakness of the will it may be said that the resulting action is guided by some motive and in so far understood as an intentional action, but the motive itself cannot

28 *Deontology*, p. 194.

29 Bowring, X, p.80.

30 Bowring, IX, p.7.

31 *First Principles Preparatory*, p. 234.

be rationally justified, and is in so far not understandable, outside the scope of an intentional description of experience. Even if Bentham did not actually disentangle all these different issues, their distinction helps us to make sense of most of what he says about the expository principle of utility.

5. THE PROOF OF THE DEONTOLOGICAL PRINCIPLE

There are two formulations of the proof that Bentham gives for the deontological principle, both of Kantian type.

The first occurs at the end of the first chapter of *An Introduction to the Principles of Morals and Legislation*. Bentham states two main theses. One of them is negative and I shall consider it first. The principle of utility, it is said, is not "susceptible of any direct proof" "for that which is used to prove everything 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"³². It is certainly open to discussion how the first principles have to be demonstrated, or agreed upon, and if such a demonstration deserves the appellation "direct proof". It is, however, clear enough that the principle of utility is not deducible from other principles in a chain of reasoning as less general principles of action may be derived from the principle of utility.

The second thesis is a positive one: the principle of utility is a necessary condition for certain moral words, such as "ought", "right" and "wrong", having meaning. In a more Kantian paraphrase one would say that such a principle is a condition of the possibility of thinking about the morality of action, i.e., a transcendental condition of both morality and the thought about it. Bentham begins with a bold assertion of the thesis:

"of an action that is conformable to the principle of utility, one may always say either that it is one that ought to be done, or at least that it is not one that ought not to be done; at least that it is not wrong it should be done; that it is a right action; at least that it is not a wrong action. When thus interpreted, the words ought, and right and wrong, and others of that stamp have a meaning: when otherwise, they have none"³³.

After this Bentham derives two corollaries of the thesis, which, if independently demonstrated, would indeed prove the main thesis. Firstly, it is impossible, without contradiction, to deny the correctness of the principle: "Has the rectitude of this principle ever been formally contested? It should seem that it had by those who have not known what

³² *An Introduction*, p. 13.

³³ *Ibid.*

they have been meaning". Secondly, it is impossible to use moral words when thinking about what ought to be done or even when discussing the principle of utility without implying the principle itself "By the natural constitution of the human frame, on most occasions of their lives, men in general embrace this principle, without thinking of it: If not for the ordering of their own actions, yet for the trying of their own actions, as well as of those of other men". "When a man attempts to combat the principle of utility, it is with reasons drawn, without his being aware of it, from that very principle itself. His arguments, if they prove anything, prove not the principle to be wrong, but that according to the applications he supposes to be made of it, it is misapplied. Is it possible for a man to move the earth? Yes, but he must first find out another earth to stand upon". The last reference to the laws of mechanics signifies that the only way to remove some application of the principle of utility is to use the same principle again differently. Neither can the principle of utility be empirically based like laws of nature, nor can it be displaced by another first principle.

Bentham imagines finally a discussion with a man disposed not to relish the principle. I shall transcribe the more relevant stretch of the argument:

- "If he is inclined to think that his own approbation or disapprobation, annexed to the idea of an act, without any regard to its consequences, is a sufficient foundation for him to judge an act upon, let him ask himself whether his sentiment is to be a standard of right and wrong, with respect to every other man, or whether every man's sentiment has the same privilege of being a standard to itself?
- In the first case, let him ask whether his principle is not despotic, and hostile to the rest of human race?
- In the second case, whether it is not anarchical, and whether at this rate there are not as many different standards of right and wrong as there are men and whether even to the same man, the same thing, which is right today, may not (without the least change in its nature) be wrong tomorrow? and whether the same thing is not right and wrong in the same place and at the same time? and in either case, whether all argument is not at an end? and whether, when two men have said, "I like this" and "I don't like it", they can (upon such a principle) have anything more to say?"

It has been noticed by Ross Harrison³⁴ that this argument strongly resembles Wittgenstein's argument against the possibility of a private language when he says: "But in the present case I have no criterion of

34 "The Only Possible Morality", *The Aristotelian Society*, Supp. Vol. 4, 1976, p. 29.

correctness. One would like to say: whatever is going to seem right to me is right³⁵. In other words, whoever uses “right” and “wrong” whenever he has a sentiment of approbation or disapprobation does not know the meaning of the rules of “right” and “wrong”.

Unfortunately Bentham does little more than to show that the standard of right and wrong in morals must be common. This is, I take it, transcendental enough.

The second formulation of the proof of the deontological principle is a kind of normative social choice theory. We find it in the “Principles of International Law”, edited by Bowring from MSS. dated 1786-89; “The end of the conduct which a sovereign ought to observe relative to his own subjects - the end of the internal laws of a society - ought to be the greatest happiness of the society concerned. This is the end which individuals will unite in approving, if they approve of any”³⁶. This is an ideal social contract theory.

A variant of the same kind of proof is to be found in the “Constitutional Code Rationale”: “In the eyes of every impartial arbiter, writing in the character of legislator, and having exactly the same regard to the happiness of every member of the Community in question as for that of every other, the greatest happiness of the greatest number of the members of that same community cannot but be recognised in the character of the right and proper, and sole right and proper, end of government”³⁷. Can the transcendental, the ideal social contract and the impartial arbiter theories be reduced to the same underlying theory or shall we choose between them? Do they imply utilitarianism? These are very much actual and difficult questions, which I cannot hope to elucidate in this paper. But they bring Bentham closer to Kant than usually admitted.

6. HOW BOTH MILL AND KANT MISREPRESENTED THE PROOF OF UTILITARISM

When Mill says that “no reason can be given why the general happiness is desirable, except that each person, so far as he believes it to be attainable, desires his own happiness”³⁸ he claims to derive the deontological from the expositor sense of the principle of utility. This presupposes a monumental misunderstanding of Bentham’s proofs of the two different senses of the principle of utility. Mill seems unaware of the importance of the distinction itself.

Similarly when Kant criticises the principle of happiness, he presupposes that it is only susceptible of empirical proof. He argues from the

35 *Philosophische Untersuchungen*, 258 (Anscombe’s translation).

36 Bowring, II, p. 537.

37 *First Principles Preparatory*, p. 235.

38 *Utilitarianism, Collected Works*, X, p. 234.

statement- that “only experience can teach us what gives us pleasure”, which Bentham would accept, to the statement that “every apparent reasoning *a priori* is here nothing else than experience raised to generality by induction”³⁹, which ignores entirely Bentham’s proof of the deontological principle, which does not depend on experience, but on the meaning of the moral words or, as Hare would say more pointedly, the logical properties of the moral concepts. The three versions of the proof are instead rather Kantian and are left untouched by the criticism. So it is clear that both Mill and Kant misrepresented Bentham’s proof.

The empirical principle of happiness that according to Kant results from “nothing more than experience elevated through induction to universality” is in Kant’s words a material practical principle⁴⁰ based on subjective ends and on the causal knowledge obtained through induction that teaches us how to achieve such ends. It is something else than Bentham’s deontological principle of utility. But Kant himself recognizes that the application of his own formal principle of the categorical imperative – and the same holds for Bentham’s deontological principle – depends on it. He says: “the natural end, that every human being has, is his own happiness. Now humanity could well survive, if nobody would contribute nothing to the happiness of others without however retiring intentionally anything from them; but this is only a negative and not positive attitude towards humanity as end in itself, if each one would not care to promote the ends of others as much as he can. This is so, because of the subject, who is himself end in itself, whose ends must, if such an idea has to have all its effect upon me, be also mine as much as possible”⁴¹. Here Kant comes as close to Bentham as ever. He distinguishes the natural empirical – Bentham would call it “expositive” – principle of happiness from a rational principle that could be called the deontological principle of happiness, which Kant conceives as an application of the categorical imperative. The sole difference seems to be that this is only an application of the categorical imperative as the only *a priori* principle of morality – Kant is speaking of not necessary duties against others – and not the whole of it.

7. CAN UTILITARIANISM GATHER THE DISTINCTION OF KANT BETWEEN AND “END IN ITSELF” AND “OTHER ENDS”?

Is it that we can also say, as did Hare, that the formal element of utilitarianism “needs only to be rephrased in order to come extremely close to Kant; there is a very close relation between Bentham’s ‘Everybody to

39 *Metaphysik der Sitten, Rechtslehre*, AB 9 (Akademie-Ausgabe 215-216).

40 *Grundlegung zur Metaphysik der Sitten*, BA 64 (Akademie-Ausgabe 427)

41 *Grundlegung zur Metaphysik der Sitten*, BA 69 (Akademie-Verlag 430).

count for one, nobody for more than one' (*ap.* Mill, 1861: ch. 5 *s.f.*) and Kant's 'Act only on that maxim which you can at the same time will that it should become a universal law' (1785:52)⁴².

The critics of utilitarianism defend generally that the similarity between both formulae is but superficial. "Everybody to count for one" simply means that each portion of pleasure or of pain registers but once in the calculation of happiness, with no respect to the human being. It does not avoid the sacrifice of the individual against his own will as long as this enlarges utility. It would justify killing indiscriminately the civil population in war, as a way of abbreviating it, or kill an innocent negro, as in "La Putain Respecteuse" of Sartre, to avoid the racial uprising which would provoke many deaths. Even if such consequences could be avoided in the name of collective well-being, it would be the case that some discrimination and some coercion of a minority of slaves for example could be justified if it contributed for the increase of general well-being. Utilitarianism could not ever guarantee human rights. Mill, in an effort to justify them, would have to seek another basis. To say this in Hart's words: "The utilities which, according to Mill, are the stuff of those universal rights to which all individuals are entitled, are forms of the individual good of those who have such rights. They are the essentials of individual human well being and things no individual human being can possibly go without. They are identified quite independently of general utility as if the criterion was to do exclusively with individual good not general utility"⁴³.

Some utilitarians tried to avoid some of these consequences, recognising that, on certain occasions, it is not consonant to utility to consult utility, to say it in words which Bentham considered full of meaning but unacceptable⁴⁴. Mill defended that the general utility is better served by those who are more concerned with virtue than with the calculus of utility. Society would mostly benefit, if it cultivated such dispositions so that the greater part of us reacted many times in a non-utilitarian way and allowed a few to become heroes and saints. Such a version of utilitarianism is in part self-effacing. It is not self-defeating, because the rule, according to which we should forget utilitarianism, is still an utilitarian rule⁴⁵.

Hare attempts to reach similar conclusions by being *rule utilitarian* at the intuitive level of moral thinking and, in this way, assuring conformity in practically everything with the dominant opinion, and act *utilitarian*

42 *Moral Thinking*, pp. 4-5.

43 H. L. Hart, "Natural Rights: Bentham and John Stuart Mill" in *Essays on Bentham*, Oxford, Clarendon Press, 1982, p. 96.

44 *An Introduction*, p. 14 n.

45 See John Stuart Mill, *Utilitarianism*, p. 235 ff.; on self-effacing and self-defeating moral theories: Derek Parfit, *Reasons and Persons*, Oxford, Clarendon Press, 1984, p. 485.

at the critical level, at which one selects the rules which should be used at the intuitive level and arbitrates between them in cases of conflict. As we are not archangels, we cannot, in conflict situations, know everything that is relevant and come to decision about it in time; so, in these situations, we should follow only our well formed intuitions, that is to say, behave as rule utilitarian. Hare's theory is not self effacing because the rules of rule utilitarianism are determined at the critical level. We should be sufficiently self-critical to allow that rule *utilitarianism* governs our habits and intuitions and also leads our actions in the majority of occasions. The difficulty arises because rule utilitarianism is ambiguous in conflict situations and giving up thinking critically is generally bad and, in such cases, even worse.

These answers make the difference between utilitarianism and the Kantian ethics of duty in practice less relevant, but do not keep away the underlying objection. The objection is that utilitarianism cannot but sustain the thesis that only pleasures i.e., a person's desires or preferences count, although only once, and, therefore, the persons themselves do not count. For Kant instead, only individual persons "count" in the sense that only they are "ends in itself", which necessarily govern the will, and therefore, are not "at the service of the will"; whilst other ends are "subjective ends", ends for one or more persons and therefore relative: their existence as a result of our action has value *for us*. For Kant, the existence of each individual person has in itself an absolute value⁴⁶.

Can then the formula "everybody to count for one" mean that no-one has greater value as a person than any other person? Can utilitarianism accept such a theory? If we depart from an *a priori* foundation of utilitarianism on the basis of the meaning of moral concepts, the aforementioned question becomes a question of knowing if the sentences containing moral words imply, due to the meaning of these words, that people who consequently should behave in a way recommended in these sentences, have a value which can not be reduced to other values which take place by their actions. To sum up: is it necessarily good or desirable to exist as a person as a necessary condition of something being good or desirable?

Hare goes to the point of saying that "I is not wholly a descriptive word but in part prescription in identifying myself with some other person either actually or hypothetically, I identify with his prescriptions. In plainer terms, to think of the person who is about to go to the dentist as myself, is to have now the preference that he should would not suffer as I believe he is going to suffer⁴⁷. In short : "To become moral is, first of all, to contemplate the hypothetical situation in which are actually going

46 *Grundlegung zur Methaphysik der Sitten*, 1785, p. 64; see also pp. 65, 77, 78.

47 *Moral Thinking*, p. 98-97.

to be states of another person would be states of oneself, and thus to acquire a hypothetical concern for the satisfaction of the preferences of oneself in that hypothetical situation; and then because of universality, to find oneself constrained (unless one takes the amoral escape route) to turn this merely hypothetical concern into an actual concern for the satisfaction of the preferences of actual other person. In plainer terms, morality requires us to argue: since if I were going forthwith to have the preferences which he actually has, I must now prescribe that they should be satisfied, and since morality admits no relevant differences between 'I' and 'he', I am bound, unless I become an amoral, to prescribe that they be satisfied. This prescription would have to compete with others, but it is enough to have secured a place in the competition. And what establishes the truth of the first 'since'-clause is the implicit prescriptivity of the word 'I'⁴⁸.

I think, Hare is right when he underlines that sentences of "duty" imply that I will accept them if I put myself in place of any other person or, what will eventually result in the same process of thought, if the preferences of others are treated as if they were mine. It is because of this that these sentences are universal. But this means that another person's preferences are morally relevant only because they are as if they were mine. It is this that is meant by taking the place of another person and this constitutes a valued difference in relation to the mere preferences of others. This does not imply that my preferences are better than yours, but that yours as much as my preferences are self-preferential, containing, therefore, a right to relevance in taking a decision. I could equally say that, taking your position, I deal with my preferences as if they were yours, which is a necessary condition in order to place them on the same level of your own preferences as a basis of a decision in your position, that is to say, in the position which you hold when you decide.

In the same line of thought, Tugendhat uses Heideggers' term "existence" when he argues that a person has to want to be or "exist" before he can want to be anything else. To want to be a person amongst all other persons includes wanting that the others "exist". In other words: to become moral is, in the first place, to have respect for oneself or to love oneself and others before having respect or loving them for one reason or another⁴⁹. The rule of mutual respect would derive analytically from any possible ethics, including utilitarianism..

48 *Ibid.*, p.223.

49 Ernst Tugendhat, *Probleme der Ethik*, Stuttgart, Reclam, 1984, p. 133 ff., 160 ff.