

ILL: 108750453



ILLiad TN: 770824



Borrower: IDU

Ship via: Odyssey

Service Level: Regular

Ariel:

Email: ill@depauw.edu

Fax: 765-658-4017

Odyssey: 163.120.6.25

Lending String:

*AFU,BDR,BIC,CDS,CFI,CSA,CSU,DKC,ESR,EZC,FXG,FXN,GPM,GZO,GZP

ISSN: 0019-0365

OCLC: 1876622

ILL - AFU

UNIVERSITY OF ARKANSAS
365 N MCILROY AVE
FAYETTEVILLE AR 72701-4002

RETURN POSTAGE GUARANTEED

LIBRARY MAIL

DEPAUW UNIVERSITY
ROY O WEST LIBRARY - ILL
11 EAST LARABEE STREET
GREENCASTLE IN 46135

Date: 9/17/2013 5:13:40 PM

Call #: **B1 .I2**

Location: **MAIN**

Volume: **22**

Issue:

Year: **12 1982**

Pages: **241-254**

Journal Title: International philosophical quarterly ;
IPQ

Article Author: PRIMORAC, IGOR

Article Title: PUNISHMENT AND UTILITARIANISM.

Notice: This material may be protected by Copyright
Law (Title 17 U.S. Code)

Charge

Maxcost: **25.00IFM**

Patron: **Anderson, Jeremy**

Initials: _____

Shelf: _____ Per: _____

Sort: _____ ILL: _____

Bad Cite: _____

Years checked _____

Table of Contents / Index _____

Utilitarianism and Punishment*

Igor Primorac

THE IMPORTANCE OF utilitarianism (= consequentialism, teleological ethics) hardly needs to be emphasized. To be sure, it is somewhat far-fetched to claim that it is "the most important of all ethical theories";¹ but it certainly is one of the most influential theories in the whole history of moral philosophy. In addition to that, it has kept a central place in the discussions in ethics to this very day.

One of the issues that have always been in the center of controversies concerning utilitarianism is the tenability of the utilitarian theory of punishment. There are two reasons for this. Firstly, utilitarianism is meant to be a social ethics *par excellence*: it does not confine itself to the sphere of individual moral choice and action, but encompasses social action, social relations, and institutions of all kinds as well. And punishment is one of the basic institutions in every political society. Secondly, the problem of justice is traditionally considered one of the most intractable for utilitarianism; and one of the main fields of application of the concept of justice is the field of criminal law. Hence the great importance that has been accorded the utilitarian doctrine of punishment both by critics of utilitarianism and by its defenders. This doctrine is but an application of utilitarianism as a general ethical theory to a specific issue in moral and legal philosophy. Accordingly, a refutation of the utilitarian theory of punishment would not amount to a refutation of utilitarianism in general, nor would a vindication of the former be equivalent to a vindication of the latter; but a well-grounded critical assessment of the utilitarian theory of punishment would clearly be an important step toward reaching a critical view of merits and demerits of utilitarianism in general.

Now utilitarianism as a general ethical theory has several features that account for its wide influence. There is certain immediate plausibility to the view that when it comes to evaluating our actions, the main consideration should be what consequences they have, together with the accompanying belief that it is only this consequentialist approach that will enable us to reason about moral issues in an objective, clear, and precise way. Other approaches leave the door open to various kinds of dogmatism, subjectivism, or sentimentalism, and are bound to result in all sorts of confusion. Further, the consequentialist approach seems to open up a possibility of constructing a theory with a rather simple structure, which will encompass the whole

*Paper read to the staff of the Department of Philosophy of the Hebrew University of Jerusalem on March 2, 1982, and the Department of Philosophy, University of Haifa, on May 10, 1982.

¹D. Lyons, *In the Interest of the Governed: A Study in Bentham's Philosophy of Utility and Law* (Oxford: Oxford Univ. Press, 1973), p. 3.

of moral experience in a system based upon a single principle, give a uniform interpretation to all *prima facie* diverse moral concepts, rules, values, and provide us with a universal method of settling moral disputes.

All these putative advantages of utilitarianism in general have been thought to accrue to the utilitarian theory of punishment; this goes a long way toward accounting for its wide acceptance in modern moral and legal philosophy. In our century it has long been almost universally accepted, often as a theory to which there is no intellectually and morally respectable alternative. The dominant view was the one that had been formulated already in Plato's *Protagoras*:

In punishing wrongdoers, no one concentrates on the fact that a man has done wrong in the past, or punishes him on that account, unless taking blind vengeance like a beast. No, punishment is not inflicted by a rational man for the sake of the crime that has been committed—after all one cannot undo what is past—but for the sake of the future, to prevent either the same man or, by the spectacle of his punishment, someone else, from doing wrong again.²

In recent years there has been something of a renaissance of the very same view of punishment which Plato saw in such a disagreeable light; but the utilitarian theory is still going strong, and has many adherents among philosophers and non-philosophers alike. For the purposes of this paper the quotation from Plato gives enough of an idea of the theory itself: of its forward-looking orientation, its appeal to various desirable consequences of punishment as its sole justification. There is no need to detail here the rather obvious different ways of preventing both the actual offender and potential offenders among the public from breaking the law in the future, which a fully developed utilitarian theory of punishment would list and systematize.³ In this paper I propose to present some arguments against the utilitarian view of punishment (section I), to point out some basic characteristics of utilitarianism that come to the fore in this connection (section II), and to draw some methodological lessons (section III).

I

Among various moral concepts that undergo a uniform, consequentialist interpretation within the utilitarian moral and legal philosophy is the idea of justice. According to a frequently quoted formulation of Bentham's, justice "is an imaginary personage . . . whose dictates are the dictates of utility, applied to certain particular cases . . . nothing more than an imaginary instrument, employed to forward on certain occasions, and by certain means, the purposes of benevolence."⁴ Consequently, within a utilitarian theory there is no need, and no room, for the idea of desert, which is logically connected with that of justice, so that within the utilitarian

² Plato, *Protagoras*, 324ab, trans. by W. K. C. Guthrie, in *The Collected Dialogues of Plato*, ed. by E. Hamilton and H. Cairns (Princeton: Princeton Univ. Press, 1973), p. 321.

³ The most thoroughly worked out formulation of the utilitarian view of punishment in philosophical literature is Bentham's theory, presented in *An Introduction to the Principles of Morals and Legislation*, *Theory of Legislation*, and *Principles of Penal Law*.

⁴ J. Bentham, *An Introduction to the Principles of Morals and Legislation*, ed. by W. Harrison (Oxford: Basil Blackwell, 1960), pp. 240–241 n. 2.

perspective "no one can deserve punishment; it can merely be right that he should be punished."⁵ Here is the root of the difficulties that plague utilitarianism in connection with the problem of punishment.

1) The alternative, retributive theory of punishment has been criticized, *inter alia*, for not leaving room for *mercy, pardon, forgiveness*. In another context I have shown that this argument works only against some versions of retributivism, but not against the retributive theory as such.⁶ Here I want to point out that it is precisely the utilitarian theory that is open to this objection. For it makes sense to speak of our mercy toward an offender, of our pardoning him or forgiving him his misdeed, only if we reduce the punishment which he has *deserved* by his deed and which is consequently *just*, or repeal it altogether. Acts of mercy, forgiveness, pardoning, are logically possible only against the background of a conflict of two principles: the principle of just, deserved punishment, and the principle of mercy and forgiveness. But within the utilitarian approach such a conflict cannot take place, for "desert" and "justice" have no standing of their own, independent of utility; the latter is the exclusive criterion for all our choices. So a utilitarian will not, and cannot, speak the stern language of justice—but precisely on account of this, he cannot speak the assuaging language of mercy and forgiveness either.⁷

2) From a non-utilitarian perspective, one of the factors to be taken into account when deciding on the measure of punishment is the degree of *provocation* under which the offense was committed. Other things being equal, the greater the provocation, the lesser the guilt, and consequently the punishment. This is the view that is accepted in practice in criminal courts. For a utilitarian, however, precisely the opposite holds true: the greater the provocation, the more severe the punishment ought to be. In other words, the very same factor that is normally taken for an extenuating circumstance, within the utilitarian theory becomes an aggravating one. For, if prevention of offenses is the main objective of punishment and its basic justification, then, as E. Westermarck put it,

...the heaviest punishment should be threatened where the strongest motive is needed to restrain. Consequently, an injury committed under great temptation, or in a passion, should be punished with particular severity; whereas a crime like patricide might be treated with more indulgence than other kinds of homicide, owing to the restraining influence of filial affection. Could the moral consciousness approve of this?⁸

3) This is not the only difficulty that the utilitarian theory meets with in connection with the responsibility of offenders. Another one, even more compromising, arises when the question of punishing the *mentally ill* comes up. It is generally held that offenders proven to be mentally ill ought not to be punished—that the state should react to their offenses in some other way, by compulsory hospitalization, for instance. Could a utilitarian subscribe to this view as well? Bentham thought that he

⁵ J. Plamenatz, *The English Utilitarians*, 2d ed. (Oxford: Basil Blackwell, 1966), p. 80.

⁶ I. Primorac, *Prestup i kazna* [Offense and Punishment, in Serbo-Croatian] (Belgrade: Mladost, 1978), pp. 30–35.

⁷ For a more detailed discussion of this issue, see A. Smart, "Mercy," in H. B. Acton (ed.), *The Philosophy of Punishment* (London: Macmillan, 1969).

⁸ E. Westermarck, *The Origin and Development of the Moral Ideas*, 2d ed. (London: Macmillan, 1912), I, 83.

could. So why not punish mentally ill offenders? Not on account of the lack of guilt or desert on their part, on account of injustice that such punishment would consequently involve; from the utilitarian point of view this is irrelevant. The reason Bentham gave was: such punishment could not be efficient. For the deterrent efficiency of punishment—and human nature being what it is, it is mainly by deterring people that we prevent them from offending against the law—presupposes the very rationality and ability to control one's actions which the mentally ill lack.⁹

But this is not a convincing reason. Of course we cannot hope to influence the behavior of mentally ill persons by threatening them with punishment if they break the law. But it does not follow from this that punishing them would have to be inefficient. For it could have a strong deterring effect on *normal* potential offenders. If the criminal law provided for mentally ill offenders to be punished just as normal ones, the perspective a normal potential offender would have to face would be much graver than the one he faces now in one important respect: he could entertain no hope of getting himself declared not responsible for his action and evading punishment for it by *simulating* mental illness.¹⁰

4) So far we have seen that it follows from the utilitarian theory of punishment that under certain circumstances it would be morally justified to mete out punishments that are considerably more or considerably less severe than those which would be deserved and just in view of the offenses committed (depending on the degree of provocation under which they were committed), and to punish those whose punishment is always undeserved and hence unjust (mentally ill offenders). On the other hand, it also follows from the theory that sometimes *the guilty should not be punished at all*. From the utilitarian point of view it is consequences, and consequences only, that constitute good reason for punishing or abstaining from punishment; desert and justice do not count in their own right. Punishment is an evil which a utilitarian considers morally justified only when it is the means for attaining a greater good—which is usually the case, but not always. Sometimes the opposite is to be expected: punishing the offender would have worse consequences than letting him get away with his misdeed; punishment would prove unprofitable. And according to utilitarianism, every unprofitable punishment is *ipso facto* morally unjustified: "It is cruel to expose even the guilty to useless sufferings," said Bentham in this connection.¹¹

When would punishing an offender be unprofitable? According to Bentham, in cases such as the following: when his punishment would cause such a great displeasure of the public, or a foreign power, that in the final account it would have worse consequences than abstaining from punishing him; or when an offender, if not punished, could render a service to the community which outweighs the utility of his punishment. In such cases punishing the guilty would "cost society too dear" and therefore would be morally unjustified.¹² From a non-utilitarian point of view, on the other hand, a criminal law system which would be guided by calculation of

⁹ J. Bentham, op. cit., p. 284.

¹⁰ Cf. H. L. A. Hart, *Punishment and Responsibility* (Oxford: Oxford Univ. Press, 1973), pp. 19–20, 40–44.

¹¹ J. Bentham, *Theory of Legislation*, trans. from the French of E. Dumont by R. Hildreth (London: Trübner, 1876), p. 345.

¹² Ibid., p. 324.

What then are we to think of the proposal that the life of a condemned criminal should be spared if he agrees to let dangerous experiments be carried out on him in order that the doctors may gain new information of value to the commonwealth, and is fortunate enough to survive? A court of justice would dismiss with contempt any medical institution which made such a proposal; for justice ceases to be justice if it can be bought at a price.¹³

Another telling example of this kind of utilitarian economy is the one Bentham recounted from a report published in *Lloyd's Evening Post* in 1776:

According to the utilitarian theory of punishment, whenever this would be the alternative with the best possible consequences under the circumstances, a feigned

¹⁴ J. Bentham, *An Introduction to the Principles of Morals and Legislation*, p. 303.

1962), I, 398.

¹⁶ Ibid., pp. 398–399 n.

punishment should be staged in place of a real one, and the public should be put in the position analogous to the position of the innocent Hottentots in this report.

6) The arguments on non-punishment of the guilty demonstrate how the utilitarian theory of punishment would in certain circumstances transform the criminal law system into a market for trading in justice, or a stage on which instead of doing justice, shows of its being done would be performed. On the other hand, under different circumstances still, the same theory would imply that it is morally justified to "*punish*" the innocent. This is the conclusion a utilitarian would have to draw whenever he found himself in a situation in which the best consequences would ensue not from acquitting, but from "punishing" an innocent man. Say, there is a country where, due to incompetence of the police and the courts, criminal behavior has become widespread, and popular trust in the legal order has been seriously undermined. The only expeditious and efficient way out of this situation, and of preventing many offenses from being committed, as they assuredly will be if the things remain the way they are, is finally to punish an offender. That punishment, together with the proper publicity, would have a strong preventive effect on a large number of potential offenders, and would help a great deal toward restoring popular confidence in the ability of the legal order and the police to provide necessary protection. However, precisely because the police are so incompetent, the opportunity never arises. Once again they have got the wrong man—the man innocent of the crime he is charged with, that is. Now his acquittal would have good consequences as far as he is concerned, but an opportunity to ensure consequences that are far more desirable from the standpoint of the common good would be lost. On the other hand, to "punish" him, his innocence notwithstanding, would have bad consequences for him, but overwhelmingly good consequences in terms of the common good. From the utilitarian point of view, it would be morally unjustified and wrong to do the former, and morally justified and even obligatory to do the latter.

7) I think that the case against the utilitarian theory of punishment can be strengthened at this point by adding an argument complementary to the preceding one, which might be termed the argument of *self-sacrifice of the innocent*. From Kant, who seems to have been the first to suggest the "punishment"-of-the-innocent argument,¹⁷ onwards, critics of utilitarianism have always viewed the situation described in the argument from the point of view of the *judge*, pointing out the unacceptable implication of utilitarianism as to what the judge would be obligated to do. But the situation could be profitably examined from another angle—that of the *innocent man accused*. If the circumstances are such that "punishing" an innocent man will have the best possible effects and, therefore, according to utilitarianism, is the morally right thing to do, then this provides the moral orientation not only to the judge, but to all those on whose conduct the outcome depends. And the accused is, obviously, one of them: in every such case, cooperation on the part of the accused can only make the trial more convincing and thereby render the positive effects expected from it and from the "punishment" which is to ensue—the effects that provide the moral justification of such a course of action—much more likely than they otherwise would be. Thus the same premises which, according to the utilitarian view of punishment, make it incumbent upon the judge to pronounce an

¹⁷I. Kant, loc. cit.

innocent man guilty and sentence him, and imply that it would be morally wrong for him to act otherwise, also make it incumbent upon the innocent man to cooperate in passing and carrying out the sentence. Regardless of the fact that he is innocent, and that he knows that the judge is aware of his innocence, he ought to collaborate assiduously in his own condemnation and "punishment" by "confessing" to the false charge and perhaps by pretending to be repentant as well. From the utilitarian standpoint, any insistence on his own innocence and any attempt to change the outcome of the trial would be morally impermissible. The only morally right thing to do for him under the circumstances is to sacrifice himself for the common good.¹⁸

II

The arguments presented in the preceding section tend to show the untenability of the utilitarian theory of punishment; at the same time, they point toward some basic features of utilitarianism in general, which are the source of some of the main difficulties this kind of ethics meets with. Let me spell them out very briefly.

When it comes to punishment, the facts of desert, of guilt or innocence, have no intrinsic moral significance for utilitarians; it is the consequences of punishment that count, and nothing else. This is but a special case of a basic attitude of utilitarianism toward *the past*. As the consequences of our actions are to be considered the sole criterion of their morality, the past as a whole has no intrinsic moral weight. To be sure, the past can be relevant, morally speaking, but only extrinsically, via the future. This exclusive orientation of utilitarianism toward the future gives rise to the well-known problems it has to face with respect to duties pertaining to relations based on facts in the past: duties of keeping promises, gratitude, loyalty to parents or friends, and such like.¹⁹

This attitude of utilitarianism toward the past and the future is also manifest in the utilitarian view of *praise* and *blame*. If we wanted to place punishment into a wider context, we would say that it is but a special case of blame, reprobation, criticism. (Some philosophers find the very essence of punishment and its basic justification in its function as a moral condemnation of the offense.²⁰) A non-utilitarian would think of this as a consideration against the utilitarian theory of punishment: he would say that praise and blame are obviously connected to desert and guilt, not only logically, but from the moral point of view as well, so that the fact that punishment is a special kind of blame suggests a retributive account of its moral basis. In reply to this, a utilitarian would refer to his own theory of praise and blame, the basic tenet of which is that praise and blame are also *actions*, so that when praising

¹⁸I. Primorac, "Utilitarianism and Self-Sacrifice of the Innocent," *Analysis*, 38 (1978), 194-199.

¹⁹The nature and scope of this obliviousness of utilitarianism to the moral significance of the past is best seen in the once famous illustration Godwin gave in the course of an argument on the moral impropriety of gratitude: if I had to choose between saving Archbishop Fénelon from his palace in flames and saving his chambermaid, given the very different importance of the two for the common good, I should save the Archbishop and leave the chambermaid to the flames—and the fact that the latter happened to be my mother should not matter, morally speaking (W. Godwin, *Enquiry Concerning Political Justice*, Bk. II, Ch. II).

²⁰The most important theory developed along these lines is the one presented by A. C. Ewing in *The Morality of Punishment* (Montclair, N.J.: Patterson Smith), 1970.

and blaming, as when doing anything else, we have to look to the future and go solely by the consequences. Now *this* kind of praise and blame will very often be quite different from the praise and blame non-utilitarians are accustomed to—and also from the praise and blame a beginner in utilitarianism might practice: sometimes the actions that are wrong from the utilitarian point of view itself will be praised, and *vice versa*.

For example, a man near Berchtesgaden in 1938 might have jumped into a river and rescued a drowning man, only to find that it was Hitler. He did the wrong thing, for he would have saved the world a lot of trouble if he had left Hitler below the surface. On the other hand his motive, his desire to save life, was one which we approve of people having: in general, though not in this case, the desire to save life leads to acting rightly. It is worth our while to strengthen such a desire. Not only shall we praise the action (thus expressing our approval of it) but we shall perhaps even give the man a medal, thus encouraging others to emulate it. Indeed praise itself comes to have some of the social functions of medal giving: we come to like praise for its own sake, and are thus influenced by the possibility of being given it. Praising a person is thus an important action in itself—it has significant effects. A utilitarian must therefore learn to control his acts of praise and dispraise, thus perhaps concealing his approval of an action when he thinks that the expression of such approval might have bad effects, and perhaps even praising actions of which he does not really approve.²¹

When it comes to “punishing” the innocent and their having to sacrifice themselves, or to staging a show of punishment instead of actually inflicting it, it is obvious that the good consequences of so doing, which make out the justification of such actions, could be attained only by keeping the facts of the matter secret and deceiving the public about them. Such a resort to *secrecy* is not something that would characterize a utilitarian only when dealing with punishment; it would be rather a practice he would have to adopt in general. Since one and the same action often has different consequences, depending on whether it is known to others or done in secrecy, it follows—as one of the foremost proponents of utilitarianism was the first to emphasize—that for a utilitarian in general

...it may be right to do and privately recommend, under certain circumstances, what it would not be right to advocate openly; it may be right to teach openly to one set of persons what it would be wrong to teach to others; it may be conceivably right to do, if it can be done with comparative secrecy, what it would be wrong to do in the face of the world; and even, if perfect secrecy can be reasonably expected, what it would be wrong to recommend by private advice or example.²²

In addition to this, of course, “The opinion that secrecy may render an action right which would not otherwise be so should itself be kept comparatively secret; and similarly it seems expedient that the doctrine that esoteric morality is expedient should itself be kept esoteric.”²³

So much for secrecy; as for the *lying* and *deception* involved, I would not go so far

²¹ J. J. C. Smart, *An Outline of a System of Utilitarian Ethics* (Parkville, Victoria: Melbourne Univ. Press, 1961), p. 35.

²² H. Sidgwick, *The Methods of Ethics*, 7th ed. (London: Macmillan, 1967), p. 489.

²³ *Ibid.*, p. 490.

as to maintain, as Quinton did in connection with "punishment" of the innocent, that "the peculiar outrageousness of the rejection of [the principle that only the guilty are to be punished] is a consequence, not of the brutality that such rejection might seem to permit, but of the fact that it involves a kind of lying."²⁴ From the moral point of view, surely, the injury comes first; still, the lying literally adds insult to injury, making a morally reprehensible action even more so. The same kind of lying and deception would be involved in self-sacrifice of the innocent. Another sort of lying and deception would be called for whenever the appearance of punishment was to be attained without actually punishing. Now this is not limited to the utilitarian theory of punishment either; lying, deception, and manipulation are means a utilitarian would be bound to adopt whenever they would be necessary for reaching the best possible results under the circumstances—which, all things considered, would be very often indeed.²⁵

Some of the arguments against the utilitarian theory of punishment presented in the preceding section refer to situations in which, according to the theory, innocent or irresponsible people would be sacrificed for the common good. Here we have another feature of utilitarianism in general—the tendency to sacrifice the individual for the sake of the group, and also the group for the sake of a larger group. This has to do with the *view of man* which is implicit in utilitarianism. Utilitarianism is a monistic ethical theory—a feature which, as I said, accounts for much of its appeal. It does not leave room for any moral concepts, norms, ideals, which would be really autonomous, i.e. which would not be susceptible to an interpretation in terms of the contribution to the common good. So there is no room for rights conceived in a way which would not reduce them to instruments for maximizing the good. Nor for the idea of man as a being endowed with dignity—a very special kind of value which is not comparable to other values; a being which is not to be treated as a means only; a being which cannot be taken as a unit of calculation or, as Arthur Koestler once put it, to which the rules of mathematics do not apply. Utilitarians conceive of people as so many receptacles of the good, and the good (be it pleasure, happiness, satisfaction of desires or interests, or whatever) is something to be quantified—not necessarily along the lines of Bentham's notoriously unsuccessful "moral arithmetics," but in some way nevertheless. So a utilitarian will normally come to moral decisions on the basis of calculations. These will not be different in any important respect when it comes to calculating the good of a single person and the good of two or more persons: "If it is rational for me to choose the pain of a visit to the dentist in order to prevent the pain of toothache, why is it not rational of me to choose a pain for Jones, similar to that of my visit to the dentist, if that is the only way in which I can prevent a pain, equal to that of my toothache, for Robinson?"²⁶ Now when one thinks about moral matters in this way, one is bound to end up using people in various ways, including sacrificing them. Rawls described this aspect of utilitarianism by saying that it conflates impartiality with impersonality, and does not take seriously the distinction between persons.²⁷ To which it might be added that this was

²⁴ A. M. Quinton, "On Punishment," in H. B. Acton (ed.), *The Philosophy of Punishment*, p. 56.

²⁵ For a detailed discussion of these issues see A. M. S. Piper, "Utilitarianism, Publicity, and Manipulation," *Ethics*, 88 (1977/8), 189–206; D. Locke, "Why the Utilitarians Shot President Kennedy," *Analysis*, 36 (1975/6), 153–155.

²⁶ J. J. C. Smart, *An Outline...*, p. 26.

²⁷ See J. Rawls, *A Theory of Justice* (Oxford: Oxford Univ. Press, 1972), pp. 22–33, 187–190.

something to be expected in an ethical theory which, owing to the view of man it assumes, does not and cannot take seriously persons themselves.

III

In this section I would like to make some points concerning the method of argument in moral philosophy, taking the critique of the utilitarian theory of punishment in section I as illustrative material. Out of the seven arguments presented there I shall concentrate on a single one—the one on “punishment” of the innocent. It has traditionally been held the strongest argument against the utilitarian view of punishment, and one of the most important for a critical assessment of utilitarianism in general; and rightly so. Before I proceed, let me point out that just to present the argument in a simple passage, as I have done above, is far from telling the whole story. Adherents of utilitarianism have made various attempts to defend their theory against it: they have claimed that it assumes a situation which is logically possible, but quite improbable, and that in ethics we are not supposed to do this kind of “moral acrobatics”; that the argument works against act-utilitarianism, but not against rule-utilitarianism; that it is inadmissible for purely logical reasons, in view of the fact that punishment is by definition inflicted upon the guilty, so that “punishment of the innocent” is a logical impossibility. I have dealt with all these maneuvers in some detail in another paper, where I hope to have shown that none of them is of any avail;²⁸ there is no need to traverse the same ground again. Here I would like merely to take up the question of the real weight of the argument.

For it is by no means obvious how far it takes us. A utilitarian could say at this point: Well, it has really been proved that the conclusion about “punishment” of the innocent being morally justified under certain circumstances does follow from my theory. But this, actually, does not worry me at all; for under those circumstances such “punishment” *would* indeed be morally justified. This is the stand taken by one of the foremost proponents of utilitarianism in our day, J. J. C. Smart. On this account a new term has been introduced into philosophical slang, “to outsmart,” meaning “to embrace the conclusion of one’s opponent’s *reductio ad absurdum* argument.” The dictionary from which I quote this definition goes on to give the following illustration of the usage: “They thought they had me, but I outsmarted them. I agreed that it *was* sometimes just to hang an innocent man.”²⁹

After this move on the part of the utilitarian it might seem that we have ended up in a blind alley. For on the other side of the fence we have non-utilitarians such as Alan Donagan, who claims that “however much certain judicial murders may be excused or even defended, they are *not* morally right,”³⁰ or Miss Anscombe, who says that “if someone really thinks, *in advance*, that it is an open question whether such an action as procuring the judicial execution of the innocent should be quite excluded from consideration—I do not want to argue with him; he shows a corrupt mind.”³¹

²⁸I. Primorac, “Utilitarianism and Punishment of the Innocent,” *Rivista Internazionale di Filosofia del Diritto*, 57 (1980), 582–625. In the following pages I draw on material of section VI of that paper.

²⁹D. Dennet and K. Lambert (eds.), *The Philosophical Lexicon*, 7th ed. (s.l.: 1978), p. 8.

³⁰A. Donagan, “Is There a Credible Form of Utilitarianism?” in M.D. Bayles (ed.), *Contemporary Utilitarianism* (New York: Doubleday, 1968), p. 198.

Now what kind of *reductio ad absurdum* do we have here? What is it that is being contradicted by the implication that "punishment" of the innocent is sometimes morally justified? In one of the most detailed accounts of the subject, the book on *Behavior Modification and "Punishment" of the Innocent* by George Schedler, we read that on this point utilitarianism comes into conflict with a "widely shared" moral intuition. The function of moral philosophy is to explain and systematize such intuitions, defined as feelings, beliefs, judgments about moral matters voiced by "plain men." These intuitions are "touchstones" of ethical theories. They are not to be questioned; an ethical theory is to be accepted if its implications square with these intuitions, and rejected if they contradict them.³² Here we have an old and well-known view of the nature of ethics, shared by philosophers such as Kant, Bradley, and Nicolai Hartmann; but its difficulties are known as well. How "widely" in space and time have these intuitions to be shared in order to qualify for "touchstones" of philosophical theories of morality? What are we to do about those ethical issues on which there are conflicting "widely shared" intuition? Is it plausible to do away with this difficulty by claiming that such disagreements are "resolvable in principle although not in practice because of human stubbornness and dishonesty"?³³ Finally, should philosophy opt *in principle* for the *status quo* in morals, and against all those who find the current moral views wrong, irrational, prejudiced, and undertake to change them—not only against Smart, but also Socrates, Bentham, Marx, Nietzsche, Russell?

I shall not dwell here on the merits and demerits of the method of appealing to the "common moral consciousness" in dealing with problems in moral philosophy. For whatever the outcome of discussing this issue might be, the method does not help at all with the problem of punishment in general, nor the question of "punishing" the innocent in particular. It is a fair guess that the "plain man" is not very consistent when facing many ethical issues; it is an established fact that he is not at all consistent when he has to deal with punishment. Bradley, who was a retributivist, argued in his time in the following way: "our people" (to whom, interestingly enough, he used to refer alternatively as "the vulgar") see the rationale of punishment in terms of retribution; now the utilitarian view of punishment is opposed to the retributive view and incompatible with it; *therefore* "our people" do not accept the utilitarian view.³⁴ But empirical research has shown that when confronted with casuistic questions on punishment, "our people" tend to switch from retributive to utilitarian judgments and back again, and that the same lack of basic consistency comes to the fore when more specific questions, relating to "punishment" of the innocent, have to be answered.³⁵ In the opinion of the authors of the study referred to, there was a more general lesson to be learnt here, namely "that the fundamental mistake of Mr. Bradley,—in which he is very far, indeed, from being alone,—lies in the assumption

³¹ G. E. M. Anscombe, "Modern Moral Philosophy," in J. J. Thomson and G. Dworkin (eds.), *Ethics* (New York: Harper & Row, 1968), pp. 206–207.

³² G. Schedler, *Behavior Modification and "Punishment" of the Innocent: Towards a Justification of the Institution of Legal Punishment* (Amsterdam: B. H. Grüner, 1977), pp. 11–20, 73–99.

³³ *Ibid.*, p. 15.

³⁴ F. H. Bradley, *Ethical Studies*, 2d ed. (London: Oxford Univ. Press, 1962), pp. 28–31.

³⁵ F. C. Sharp and M. C. Otto, "A Study of the Popular Attitude towards Retributive Punishment," *International Journal of Ethics*, 20 (1909/10), 341–357; "Retribution and Deterrence in the Moral Judgments of Common Sense," *ibid.*, pp. 438–453.

that any one consistently maintained standard forms the basis of the judgments of 'our people' in the matter of punishment, or, indeed, of anything else."³⁶

So let us forget about the "common moral consciousness" and turn to our own. What are we to do if we feel that "punishment" of the innocent is morally unacceptable even if it is the course of action which would have the best consequences attainable under the circumstances? What reasons can a utilitarian give to make us suppress our moral repugnance to such "punishment" in order to be able to accept wholeheartedly his theory together with all its implications? In his monograph *An Outline of a System of Utilitarian Ethics* J. J. C. Smart had this to offer:

Some writers seem to think that to refute utilitarianism it is sufficient to show that it conflicts with some of our particular moral intuitions or feelings. . . . A false analogy with which science seems to operate. In science we certainly must correct our general principles in the light of particular observations. But in ethics our particular statements are not observation reports but recommendations. It therefore seems to me that the matter is the other way round. Our general principle, resting on something so simple and natural as generalized benevolence, seems to me to be more securely founded than our particular feelings, which are subtly distorted by analogies with similar looking (but in reality totally different) types of cases, and with all sorts of hangovers from traditional and uncritical ethical thinking.³⁷

The first thing one could object here is that there does not seem to be any reason why particular moral reactions should be considered less simple or less "natural" than generalized moral feelings; secondly, it is not obvious why that which is simpler or "natural" should carry greater moral weight than that which is complex or "not natural" (which here presumably means "acquired" or "learned"). Moreover, a careful reader of Smart's study is bound to be surprised to find Smart disqualifying the method which he himself used in an earlier chapter when criticizing an alternative theory.³⁸

This last point is the most telling: it is not at all accidental that an author who, when replying to objections to his theory, explicitly rejects the method of testing ethical theories with the help of our immediate moral reactions, used the very same method when criticizing other theories. For what other method could he have used? To test an ethical theory means, first and foremost, to test what the theory proclaims to be the fundamental moral principle. A fundamental moral principle, being *fundamental*, can be neither established nor refuted from the standpoint of another *moral* principle. Being a *moral* principle, it cannot be derived from a principle which is *not moral*, nor can it be criticized for not having been established in such a way. On the other hand, anything proposed as a fundamental moral principle will invariably be very general. Almost any such principle, taken by itself, will appear quite acceptable. But we shall not be certain that we can really accept such a principle until we have a better idea of its contents. And the contents of a moral principle are its practical implications. Thus, the prescription that we strive to maximize the good and

³⁶F. C. Sharp and M. C. Otto, "Retribution and Deterrence in the Moral Judgments of Common Sense," p. 444.

³⁷J. J. C. Smart, *An Outline* . . . , pp. 40-41.

³⁸Cf. *ibid.*, p. 19. This was pointed out by Smart's critic R. T. Garner in "Some Remarks on Act Utilitarianism," *Mind*, 78 (1969), 124-128, at pp. 125-126.

minimize the bad, which is the fundamental principle of utilitarianism, is *prima facie* quite plausible. Still, we cannot know whether we can really accept it as a principle by which to make our moral choices and to judge morally actions of others, until we have examined what it would actually mean to strive to maximize the good and minimize the bad in various kinds of situations that might arise, that is, what the principle would imply we ought to do in those situations. And there seems to be no way of deciding for or against a particular implication of a principle in a specific situation except by seeing whether it is identical with, or opposed to, our own immediate moral reaction (intuition, feeling, judgment) in that situation.

This does not mean that these immediate moral reactions of ours are sacrosanct. We shall probably be prepared to correct some of them, either because they contradict some others which are more deeply ingrained, or because they are incompatible with some moral rule or principle by which we firmly stand, not willing to give it up or to modify it. But there are also bound to be those which we will not be prepared to give up; for otherwise we would have no firm ground on which to stand when deciding for or against various moral principles or various ethical theories. The final outcome of a quest for moral orientation in some particular field, or even in the sphere of moral experience as a whole, which would proceed along these lines, might be a state which Rawls has termed "reflective equilibrium."³⁹

So the method of critically examining ethical theories by confronting their implications with our immediate moral reactions is perfectly legitimate. The effectiveness of applying this method, however, will depend in each particular case on whether these reactions will be actually in keeping with what the critic expects. Consequently, the critic of utilitarianism, leaving the common morality aside, can present the argument on "punishment" of the innocent as an argument which is *ad hominem* in that it is addressed to each of us personally. Faced with this argument, everyone can envisage the situation described to see what his own immediate moral reaction would be. I imagine that some of us will feel that it would be morally right in such a situation to "punish" the innocent man; what the critic of utilitarianism presented as an argument against that theory, such persons will see merely as part of its analysis. In such cases the critic will find himself "outsmarted"—not by the utilitarian, but by his own readers. But I also believe that many will evince the conviction that "punishment" of the innocent is morally intolerable even when it would have better consequences than any other possible action. For those persons, this argument will eloquently speak against the ethical theory which morally consecrates such an injustice.

In an article written after the aforementioned monograph, J. J. C. Smart tried to turn the argument about "punishment" of the innocent against the non-utilitarian viewpoint in ethics:

Surely if it is shown that, in certain circumstances... a utilitarian ought, on his own principles, to commit a serious injustice, such as punishing an innocent man, then it seems that this *does* and *should* weaken the appeal of utilitarianism. And yet one can be made to vacillate back again. We also reflect that the serious injustice would *ex hypothesi* be the only possible alternative to an even greater total misery than would be caused by the injustice. The anti-utilitarian conclusion is a very unpalatable one too, namely that in some circumstances one should choose the greater total misery. ... It seems to me that some of the im-

³⁹ J. Rawls, *op. cit.*, pp. 20-21, 46-51.

plications of any ethical theory will be unsatisfactory. If the theory is a utilitarian one, then the possibility that one ought to commit injustice will be felt to be acutely unsatisfactory by someone who has had a normal civilized upbringing. If, on the other hand, it is a theory with deontological elements, then it will have the unsatisfactory implication that sometimes avoidable misery ought not to be avoided.⁴⁰

Now this is very true. The argument on "punishment" of the innocent strikingly shows that, no matter how much the good and the just may frequently coincide, this is by no means inevitable; there are cases when the two come into sharp conflict, so that we cannot but run counter to one of them. Therefore we have to choose—in ethical theory, and in life itself as well. One choice will proceed from the view that the good of man is the substance of morality and its supreme criterion, and that, consequently, the execution of justice is meaningless, morally and humanly unacceptable, if it conflicts with that. The other choice will express the conviction that justice is not merely an instrument of the common good, but an autonomous moral principle, whose demands ought to be carried out even when this means paying a price in human well-being, in happiness and misery.

⁴⁰J. J. C. Smart, "The Methods of Ethics and the Methods of Science," *Journal of Philosophy*, 62 (1965), 344-349, at pp. 347-348.