DOES THE DETERRENCE THEORY OF PUNISHMENT EXIST?: A RESPONSE TO NOZICK

Robert F. Ladenson

January 29, 1976
Abstract:

Among the many assaults upon widely held views in social and political philosophy to be found in Robert Nozick's *Anarchy, State, and Utopia*, is a novel criticism of the utilitarian deterrence theory of punishment. Nozick believes that this criticism is absolutely decisive, and, indeed, in his words, establishes the utilitarian deterrence theory's "non existence." The purpose of this paper is to show that Nozick's criticism rests upon a tacit crucial error about the nature of punishment. This error, while an elementary one, is evidently easy to make since not only Nozick falls prey to it but also some prominent utilitarians themselves. Recognizing the error makes possible a more careful statement of the utilitarian deterrence theory that avoids Nozick's criticism.
Does the Deterrence Theory of Punishment Exist?: A Response to Nozick

Among the many assaults upon widely held views in social and political philosophy to be found in Robert Nozick's *Anarchy, State, and Utopia*, is a novel criticism of the utilitarian deterrence theory of punishment. Nozick believes that this criticism is absolutely decisive, and, indeed, in his words, establishes the utilitarian deterrence theory's non "existence."¹ The purpose of this paper is to show that Nozick's criticism rests upon a tacit crucial error about the nature of punishment. This error, while an elementary one, is evidently easy to make since not only Nozick falls prey to it but also some prominent utilitarians themselves. Recognizing the error makes possible a more careful statement of the utilitarian deterrence theory that avoids Nozick's criticism.

Nozick states his objection thus:

"...'the (principle that the) penalty for a crime should be the minimal one necessary to deter commission of it' provides no guidance until we're told how much commission of it is to be deterred. If all commission is to be deterred, so that the crime is eliminated, the penalty will be unacceptably high. If only one instance of the crime is to be deterred, so that there is merely less of the crime than there would be with no penalty at all, the penalty will be unacceptably low and will lead to almost zero deterrence. Where in between is the goal and penalty to be set? Deterrence theorists of the utilitarian sort would suggest (something like) setting the penalty P for a crime at the least point where any penalty for the crime greater than P would lead to more additional unhappiness inflicted in punishment than would be saved to the (potential) victims of the crimes deterred by the additional increment in punishment.

This utilitarian suggestion equates the unhappiness the criminals punishment causes him with the unhappiness a crime causes its victim. It gives

¹. *Anarchy, State and Utopia*, p. xi.
these two unhappinesses the same weight in calculating a social optimum. So the utilitarian would refuse to raise the penalty for a crime, even though the greater penalty (well below any retributive upper limit) would deter more crimes, so long as it increases the unhappiness of those penalized more, even slightly, than it diminishes the unhappiness of those it saves from being victimized by the crime, and of those it deters and saves from punishment. (Will the utilitarian at least always select, between two amounts of penalty that equally maximize the total happiness, the one that minimizes the unhappiness of the victims?) Constructing counterexamples to this bizarre view is left as an exercise to the reader. 2

Nozick's complaint then is that the utilitarian deterrence theory equates the unhappiness the violator's punishment causes him with the unhappiness a violation causes its victim; and this, he feels, is totally bizarre as a conception of how to assign penalties to particular offenses. Toward the end of the above paragraph (in the sentence in parentheses) Nozick alludes briefly to one especially odd seeming implication of this conception that the following table illustrates in more detail:

<table>
<thead>
<tr>
<th>&quot;Sanction&quot; for committing offense N</th>
<th>Happiness and/or unhappiness for people who commit offense N</th>
<th>Happiness and/or unhappiness for people who do not commit offense N</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Six months in jail</td>
<td>-80</td>
<td>200</td>
<td>120</td>
</tr>
<tr>
<td>(2) Thirty days in jail</td>
<td>-50</td>
<td>170</td>
<td>120</td>
</tr>
<tr>
<td>(3) Slap on the wrist</td>
<td>180</td>
<td>-60</td>
<td>120</td>
</tr>
<tr>
<td>(4) Pat on the back</td>
<td>200</td>
<td>-80</td>
<td>120</td>
</tr>
</tbody>
</table>

2. Ibid., pp. 61-62.

3. To avoid misunderstanding, it should be stated that the above table presupposes the following five propositions to which, I assume, utilitarians are committed:

   (1) At any given time, any given person is at
The problem is that if the happiness and/or unhappiness the violator's punishment causes him is to be equated with the happiness and/or unhappiness it causes his victims (and potential victims), it appears that we are saddled with the consequence that "sanctions" (1) through (4) above are all on a par. But this hardly seems correct. Nozick maintains that the only way out if the utilitarian deterrence theory is maintained is to accord greater weight to the happiness and/or unhappiness of the victim than to that of the violator. Taking this tack, however, Nozick feels, can only, at best, mitigate the consequences of a fundamentally mistaken account. The better course, he holds,

_____________________________

some 'hedonic level' or other, positive or negative or neutral -- that is, either happy (in some degree or other) or unhappy (in some degree or other).

(2) The range of possible hedonic levels forms (or approximates) a continuum--one's hedonic level can move up or down either continuously or by very small increments.

(3) There is in principle some 'hedonic metric' such that the hedonic state of a given person at a given time can be compared with any state of himself at another time or with that of another person at any time.

(4) This 'hedonic metric' is such that cardinal numbers (positive and negative) can be assigned to hedonic levels.

(5) A hedonic arithmetic and calculus are thus possible: a person's "hedonic sum" can be (in principle) computed for any given period of time and his 'hedonic index' (that is, average hedonic level) for that time computed; and hedonic sums and indexes of groups of people can similarly be determined."

(The above list appears in Richard Henson, "Utilitarianism and the Wrongness of Killing" Philosophical Review: July, 1971)

4. Nozick believes this because, on his view, by giving more weight to the happiness and/or unhappiness of the victims than to that of the violator, the utilitarian thereby introduces considerations of desert into his theory which he is (ought to be) anxious to avoid.
is to embrace retributivism.

A utilitarian might respond to the above criticism by pointing out that the situations depicted in the case of penalties (3) and (4), though logically possible, are absolutely inconceivable in light of any reasonable conception of human nature; and thus, the entire table is of no real interest. That is, a utilitarian might hold that for any social group that has ever existed, or that we have reason to believe will ever exist, it is inconceivable that a policy of slapping the wrists (or even patting the backs) of people whose actions cause substantial harm to others could cause so much happiness for those to whom the "sanction" is applied that it would offset the unhappiness that their actions cause others. This response, though very possibly correct, is not the most interesting one that can be made, however, since it fails to reach a deeper error in Nozick's criticism, the recognition of which paves the way for a more refined statement of the utilitarian deterrence theory of punishment.

It is important to notice that on Nozick's view, Utilitarians are committed to balancing the happiness and/or unhappiness of criminals against that of their victims (and potential victims) in virtue of advocating that the following method be employed in affixing penalties to particular offenses:

Set the penalty \( P \) for a crime at the least point where any penalty for the crime greater than \( P \) would lead to more additional unhappiness inflicted in punishment than would be saved to the (potential) victims of the crimes deterred by the additional increment in punishment.

On the face of it this method appears to embody characteristically utilitarian reasoning, and, indeed, one finds it embraced by some prominent Utilitarians: for example, Richard Brandt writes,

"Let us suppose a Utilitarian is considering what should be the penalty affixed to the act of homicide. He will follow Bentham in deciding such matters by applying the principle of marginal utility to punishment, or what might be called the principle of 'least necessary punishment'. For since punishment is an evil, the Utilitarian will want to inflict it only in order to avoid a greater cost. In the case of homicide, the Utilitarian will insist on having some penalty for the offense; otherwise,
everyone is invited to indulge in homicide when it suits his ends. The utilitarian will consider successively more severe penalties, and each time he will count the loss of the punishment against the predicted gain, primarily in crime prevention because of the threat of punishment. When he arrives at a degree of severity such that the utility of the crime prevention just equals the disutility to society of the punishment itself, he will decide that the punishment is just severe enough."

Brandt refers to the method of affixing penalties to particular offenses described in the above paragraph as the principle of least necessary punishment. Since, as will be shown shortly, however, there is another approach to the problem of deciding upon penalties for specific offenses that is more appropriately termed the principle of least necessary punishment, it will be less confusing if the method described by Brandt is referred to by some other name. Accordingly it will be hereafter referred to as the Brandt method.

Nozick believes that the Brandt method is "bizarre" because it involves balancing the happiness and/or unhappiness of criminals against that of their potential victims. The method, however, has a graver defect; viz. that it rests upon the clearly false assumption that beliefs about the penalty attached to an offense, by themselves, are sufficient to deter people from committing the offense. One can see that the Brandt method rests upon this false assumption by examining it more closely. According to Brandt, the very first step in the method is to consider successively more severe penalties, each time counting the loss, from a utilitarian point of view, against the predicted gain of the threat of the penalty in virtue of its deterrent effect. But if one has good reason to believe that he will never be apprehended or, that if apprehended the penalty will not be applied, then the mere belief that a certain penalty is attached to a certain action should hardly deter one from doing it. What deters people from committing an offense then is beliefs about the penalty attached to that offense together with beliefs about the likelihood of being apprehended should one commit the offense, and beliefs about the likelihood of being punished should one be apprehended.

Accordingly, for the utilitarian there are always three distinct questions to be dealt with once it has been decided (on utilitarian grounds) that an action N is to be made punishable:

(a) What penalty should one be liable to pay for committing N?

(b) What should be the policy with respect to trying to apprehend alleged perpetrators of N?

(c) Given that an individual A has been apprehended and found guilty of doing N, should we go ahead and apply the penalty to which he is liable?

In the most important case, where punishment is an activity of the state, questions (a), (b) and (c) are generally dealt with respectively by legislators, chiefs of police, and judges. These three distinct questions, however, will always occur regardless of who does the punishing. Thus, parents, school officials, and human beings in the state of nature (as conceived by Locke and Nozick among others) must face these questions no less than civil authorities.

A utilitarian theory of punishment then involves explanations of how a utilitarian would go about answering the above three questions. When one considers this issue, however, it turns out that the Brandt method does not at all figure in utilitarian answers to either (a), (b), or (c). More important, the respective utilitarian responses to these questions in no way involve balancing the happiness and/or unhappiness of the violators of an offense against that of their potential victims, as does the Brandt method.

6. Parents, for example, must not only decide what to tell their children is the penalty to which they are liable for doing certain deeds, but also they must decide how much effort they will expend in finding out whether their children did those deeds. Besides that, in each case in which parents discover that their children have done what they were forbidden to do presents the further problem of whether to apply the penalty that they told their children beforehand they would be liable to pay for behaving as they did. In the same way, Nozick's protective associations would face similar questions once they took it upon themselves to prohibit certain activities. (See Anarchy, State and Utopia, Chs. IV and V).
Thus Nozick's criticism of the utilitarian deterrence theory of punishment is incorrect. The following discussion will establish the above claims.

Consider first how a utilitarian rule maker would go about deciding what penalty a person should be liable to pay for committing an offense N. For the sake of simplicity, only the case of a legislator in a commonwealth will be considered, though the following remarks apply, with only slight modification, to the case of any rule maker. From the legislator's standpoint the ideal situation would be where all violations of N are prevented and the unhappiness resulting from the policies responsible for this is minimal. The legislator, however, when drawing up, or amending, the penal code has little direct control over the administration of the police, whose level of professionalism and number of personnel determine, to a large extent, the degree to which people believe that they will be apprehended if they violate N. The legislator can, at some other time, consider measures whose purpose is to increase the effectiveness of the police department. Such measures, however, generally involve issues (e.g. funding for personnel and equipment) that are not directly related to the issues before the legislator in deciding what penalties to affix to certain offenses. In any case, the legislator's job is to make laws and not to run the police department on a day to day basis.

The legislator then must make decisions about what penalties to affix to violations of N without knowing very much in advance about the extent to which violators of N will be apprehended. Accordingly, the question for a utilitarian theory of punishment is 'What is the best way for the legislator to make decisions of this kind from a utilitarian point of view?' One initially plausible suggestion in this regard is that since the legislator has no direct control over the level of effectiveness of the police force at the time he draws up the penal code, and since exercising the control he has will involve considering issues that cannot be decided upon in advance, the legislator should not try to take account of the police department's probable level of effectiveness in his deliberations at all, but instead concentrate solely on the following question:

What is the least penalty $P$ sufficient to deter a rational person from committing N if the person is certain that if he commits $\bar{N}$ he will be apprehended, convicted, and thus liable to be punished?

The suggestion that legislators should ask themselves
the above question in deciding what penalties to affix to particular offenses clearly reflects the utilitarian outlook on punishment. It implies, for example, that if the certain expectation of being liable to spend five years in prison is sufficient to deter any rational person from cutting off someone else's right arm, then this penalty is preferable to any other more severe penalty (for example, cutting off the offender's own right arm or, for that matter, cutting off his left arm, or even cutting off his left thumb). Thus, attempting to affix penalties in this way can be thought of as adhering to the Doctrine of Least Necessary Punishment (hereafter LNP). Unlike the Brandt method, however, LNP does not presuppose the false assumption that beliefs about the penalty affixed to a certain action can be sufficient, by itself, to deter a rational person from performing that action. Equally important, it does not involve balancing the happiness and/or unhappiness of criminals against that of their potential victims. It seems then that Nozick's criticism does not apply to the first question ((a) above) with which a utilitarian theory of punishment must deal.

There are, however, a number of further points that must be made about LNP that are best presented in the form of replies to the following four possible objections to it.

Objection 1: LNP falsely assumes that for each offense there is a unique least punishment that suffices to deter a rational person. To the contrary, one rational person might be deterred by the threat of a stern lecture while for another it might take the threat of many years in prison.

The above objection confuses two kinds of personal characteristics - viz., those that determine whether a person will run the risk of suffering a certain penalty for violating the law, and those that determine the price a person is willing to pay for such a violation, on the assumption of certain apprehension and conviction. While rational

7. Neither this example nor anything else in this paper should be taken as implying that imprisonment is the only possible penalty on the utilitarian deterrence theory. In this regard the growing dissatisfaction in the United States with regard to current practice in the area of punishment does not reflect a rejection of the idea that the primary function of punishment is deterrence so much as it reflects acceptance of the proposition that in our society the threat of punishment does not have a very great deterrent effect.
people vary greatly with regard to the former characteristics (i.e., with regard to such things as boldness, resoluteness, liability to optimism or pessimism, and risk aversion) the degree of variance is not so great with regard to the latter characteristics. All rational people desire to avoid certain evils, such as death, pain, and loss of liberty. Moreover, while their respective attitudes toward how they would trade off the certain prospect of suffering such evils against the certain prospect of realizing various personal gains from violating the law do not coincide with great precision, there is enough coincidence to predicate a rational policy of attaching penalties to offenses upon it.

Objection 2: LNP appears to avoid Nozick's criticisms because it is formulated in terms of what would deter a rational person from violating the criminal law. In reality, however, legislators are not concerned with deterring rational persons so much as with deterring whatever persons live in their society.

This objection fails to realize that the criminal law, of necessity, presupposes at least a minimal level of rationality on the part of individuals in society. If a person was so irrational as not to be deterred from committing a crime by the prospect of thereby suffering certain evils, such as death, pain, or loss of liberty, or if a person's schedule for trading off the prospect of suffering such evils against the prospect of obtaining certain gains by doing so was inexplicable bizarre, then there could be no basis for beliefs about the kinds of penalties that would deter such a person. The "irresistible impulse" test of criminal insanity, by which one is judged criminally insane, and hence not criminally liable, if one would commit a felony knowing that a police officer was standing nearby and watching, nicely illustrates the presumption of a minimal level of rationality on the part of individuals in society inherent in the criminal law.

Objection 3: LNP presupposes that in drawing up the penal code legislators have no idea what-so-ever about the extent to which violators of various offenses will be apprehended. It would seem, however, that in some cases they can have at least a few general ideas about this matter, and that such ideas can be relevant in their deliberations about the appropriate penalties for various offenses. Suppose, for example, that the chief of police in a certain municipality testifies, in a closed session of the legislative committee amending the penal code, that in the case of certain offenses, it is so difficult to apprehend violators that even if the police force devoted itself to
nothing else but investigating alleged violations of these offenses, a significant number of violators would still go unapprehended. It seems plausible to suppose that in such cases the legislator may well consider raising the penalties for the offenses in question so as to compensate for the low probability of being apprehended.

Affixing penalties in the above manner does not mark a significant departure from the method prescribed by LNP. That is, rather than considering what is the least penalty $P$ sufficient to deter a rational person from committing an offense if the person is certain that if he commits the offense in question he will be apprehended, convicted, and thus liable to be punished, the legislator instead considers what is the least penalty $P$ sufficient to deter a rational person if the person is certain that if he commits the offense in question there is such and such a probability that he will be apprehended, convicted, and thus liable to be punished. To be sure, since in many situations the spectrum of attitudes toward risk that are allowed by reason is rather broad, a legislator's task is here more complicated: viz., it is to determine the least penalty sufficient to deter someone who, while rational, is not very risk averse. Nonetheless, affixing penalties in this latter way, no more involves application of the Brandt method, with its balancing of the happiness and/or unhappiness of criminals against that of their potential victims, than does the former method.

Objection 4: LNP presupposes that a utilitarian legislator, in drawing up the penal code, would allow judges considerable leeway in the sentencing of criminals. That is, the legislator is conceived of as trying to decide what penalties a person should be liable to pay for committing a certain offense by attempting to ascertain the least penalty $P$ sufficient to deter a rational person from committing that offense if he is certain that if he commits it he will be (there is such and such a probability that he will be) apprehended, convicted, and thus liable to be punished. One can question, however, why such leeway should be allowed to the judge. After all, assuming equal conviction rates for a given offense, would not a policy of requiring violators to pay the prescribed penalty have a greater deterrent effect than a policy of merely making them liable to pay it? By allowing judges great leeway in sentencing do we not run the risk that overly lenient judges may consistently undermine the intended deterrent effects of prescribed penalties?

The answer here is that the gains, from a utilitarian standpoint, to be derived from a policy of mandatory
sentences must be weighed alongside of the attendant losses, specifically from inflexibility in the sentencing procedures. Violations of an offense do not come uniformly packaged. That is, differences in the violator's motivation and circumstances often call, from a utilitarian standpoint, for different responses. Such differential responses, however, can only be written into the law itself in a limited way at best (e.g., the differences between the penalties for different degrees of murder). There will always be many individual cases that arise in which applying the prescribed penalty for a given offense is simply not a sensible utilitarian response, and the only way to increase the flexibility of the penal system so that it can deal with such cases is to grant judges rather broad discretionary powers in sentencing. Thus the problem faced by a utilitarian legislator is to design a penal system in which there is a satisfactory trade-off between deterrent effect and flexibility. Now it is plausible to hold that if trials are fair and judges are competent then, on the whole, a sufficiently credible deterrent threat can be maintained by merely making people liable to punishment for certain offenses rather than by requiring in all cases that they be punished. Such is the case because first, if trials are fair then all relevant evidence and testimony will tend to be presented; and second, if the judge is competent then he will be capable of intelligently deciding on the basis of that evidence, whether or not there are special features of the case at hand in virtue of which the prescribed penalty should be altered downward. Accordingly, a rational person who knows that if he commits a given offense he will be apprehended and convicted will also know that unless he can do a very good job of deceiving the judge about his motivation and circumstances then, in all probability, the judge will apply the prescribed penalty. Remember that, by hypothesis, all the evidence is likely to come out at the trial and the judge is no fool. Thus a utilitarian legislator would opt for a penal code in which violators are generally liable to pay a penalty rather than always required to pay it. Just how broad a utilitarian legislator would want to make the judge's discretionary power in sentencing, however, would vary from offense to offense.

The first question then with which a utilitarian must deal with once it has been decided that an action is to be made punishable - viz., (a) What penalty should one be liable to pay for performing it? - can be answered without balancing the happiness and/or unhappiness of violators of the law against that of their victims. It will be recalled, however, that two other questions remain for the utilitarian, namely, (b) What should be the policy with respect to trying to apprehend alleged violators of a given law, and (c)
Given that an individual has been apprehended and found guilty, should we go ahead and apply the penalty to which he is liable? Each question can be answered in short order. A utilitarian police chief will deal with question (b) by first attempting to formulate an intelligent position on the matter of how important it is (on utilitarian grounds) to prevent commissions of the offense in question, as opposed to other offenses. Given a position on this matter, he will then consider to what extent the resources of the police force can be devoted to tracking down alleged violators. Thus, in answering (b) the police chief will weigh the unhappiness to be prevented by minimizing the commission of the offense in question against the unhappiness to be prevented by minimizing the commission of other offenses. He will not employ the Brandt method, which in this case would involve weighing the respective levels of happiness and/or unhappiness of criminals and their potential victims deriving from different enforcement policies.

The third question, (c) arises after an individual has been apprehended and convicted for violating a given law; viz. "Should we now go ahead and apply the penalty to which he is liable?" Such is the issue before the sentencing judge who, as argued above, is likely to have broad discretionary powers within a utilitarian penal system. The judge's problem is a difficult one. He must decide which among the many ways of treating the offender that are possible within the constraints imposed by the penal code will maximize the net balance of happiness over unhappiness all things considered. Owing to the unique problems that tend to arise from one case to the next, there is very little of a general nature that can be said about how a utilitarian judge should make decisions of this kind. To be sure, in most cases, one among the many factors he will consider is the probable effects upon the violator of the various penalties that can be imposed upon him in light of the provisions of the penal code. But a judge's leeway in sentencing notwithstanding, the penal code places restrictions upon the extent to which he can take this factor into account. Thus, a utilitarian judge will no more deal with this question by straightforwardly trying to balance the happiness and/or unhappiness of the criminal against that of his potential victims than will the legislator and police chief in dealing with their respective questions. Rather, he will attempt to strike a satisfactory balance (from a utilitarian point of view) between a wide variety of factors, and he will carry out this attempt within constraints.
imposed by the provisions of the penal code.8

To summarize briefly, there are three separate questions with which a utilitarian must deal with once it has been decided that a certain act will be made punishable. When one considers plausible utilitarian approaches to these questions, however, it turns out that the method which Nozick, and some prominent utilitarians as well, regard as central to the utilitarian deterrence theory of punishment does not figure at all. Thus since the method which Nozick finds objectionable plays no role in the utilitarian approach to punishment his criticisms are misdirected. Contrary to Nozkck, we can not only say that the utilitarian deterrence theory of punishment exists, but that if properly understood it will be around for a long time to come.9

Robert F. Ladenson
Department of Humanities
Illinois Institute of Technology
Chicago, IL 60616

8. It might be objected that insofar as the above account of how an ideal utilitarian judge would discharge his duty in sentencing entails that he would take the pain that a punishment might cause to an offender into consideration, the account fails to rescue utilitarianism from Nozick's criticism. Such an objection, however, is mistaken. The problem posed by Nozick's criticism is not to rescue the utilitarian deterrence theory from being committed to holding that the pain caused by a punishment to the offender should be taken into account. Rather, it is to rescue the utilitarian deterrence theory from bizarre consequences (such as those brought out by the table of penalties mentioned earlier) that allegedly follow from holding that the pain of the punishment to the offender should be taken into account.

9. The author wishes to acknowledge the helpful criticisms of Gary Thrane, and Vivian Weil.