

VI.-PUNISHMENT FOR INTENTIONS

BY GERALD DWORKIN and DAVID BLUMENFELD

ONE of the basic features associated with the concept of punishment is that no person should be punished unless he is guilty of having committed some wrong act. We punish a man for attempted murder but not for wanting to murder. We expel the student who cheated on the examination but not his friend who would have but for the lack of opportunity. Even if we agree that " whosoever looketh on a woman to lust after her hath committed adultery with her already in his heart " we do not punish this man for adultery. Similarly, as to degree of punishment we distinguish between those guilty of conspiring to defraud and those who actually de fraud. The swindler receives a harsher penalty than his unsuccessful emulators. The man who through ignorance of toxicology provides his victim with less than a fatal dose of poison is treated more leniently than one who is more knowledgeable. We take these examples as data by which to test the adequacy of a number of plausible contemporary ethical theories. We shall concentrate on two examples which have the features illustrated above and attempt to show that none of these theories can account for certain of the practices which we ordinarily take to be just.

Case 1A. Mr. X plans to shoot Mr. Z. He raises his gun, fires, but at the last moment Z moves unexpectedly and X misses. Case 1B. Mr. Y plans to shoot Mr. Z. He raises his gun, fires, and hits and kills Z. Case 2. In a small South American republic, Señor X and Señor Y decide independently to assassinate the President. They both plan to fire from the same rooftop at the same time during the President's procession. Unfortunately for Señor X his alarm clock fails to ring and by the time he arrives at the scene Señor Y has already accomplished the mission. In Case 1A, X would be punished for attempted murder and receive a less severe penalty than would Y who is guilty of murder. In case 2 only Y would be punished at all. In judging the character of X and Y in Case 1 there is no morally relevant factor by which to differentiate them. It was merely a matter of chance that affected the outcome. Their success or failure at the deed is not considered when we evaluate 396

This content downloaded from 131.96.12.74 on Tue, 24 Apr 2018 20:37:41 UTC All use subject to <http://about.jstor.org/terms>

PUNISHMENT FOR INTENTIONS 397

their 'moral worth '. The considerations that this judgement involves are diverse and include the individual's intentions, desires, and dispositions. Of course, the person's past deeds are taken into account as aids in identifying the nature of his intentions and dispositions. Similarly, in Case 2 we see no difference in the moral worth of X and Y since the failure to commit a crime was not due to any attribute of the agent but to chance. Since we do not differentiate between the individuals in terms of moral worth and we do distinguish between them with respect to assigning punishment we can conclude that we do not punish solely on the basis of moral worth. Why not? One might plausibly argue that unless there is a morally relevant factor by which to differentiate between two individuals they should be punished equally. But since in the cases above their moral worth is equal there is no such factor and the punishment should be equal. It seems then that justice makes demands which run counter to our practices. To this it will be

objected that there are morally relevant factors which can serve to distinguish the individuals with respect to assigning punishment. In Case 1, Y did harm which X, if only accidentally, failed to do. In Case 2, Y did harm which lack of opportunity prevented X from doing. In each case Y did something that X did not do. While it must be admitted that there is this difference, what is lacking is an explanation of the moral relevance of the difference. Simply to assert "the deed, the deed" would seem to turn accident or lack of opportunity to the advantage of the agent. The general question we are asking can be put this way. Why do we differentiate individuals with respect to assigning punishment when our moral assessment of them is the same? Furthermore, we wish to emphasize that the question we are asking is not, "Why don't we always punish people for their bad intentions? but specifically " Why don't we punish individuals for bad intentions when the only factor preventing them from translating these intentions into actions is lack of opportunity or accident?

II

Let us consider how the act- and rule-utilitarian and the retributivist would account for our practices in Case 1. An act utilitarian claims that punishment is justified if and only if punishment serves to reform the individual or to deter similar actions. As to proportion between punishment and offence implicit in the theory is the notion that no greater punishment should be inflicted than is needed to produce the desired results,

This content downloaded from 131.96.12.74 on Tue, 24 Apr 2018 20:37:41 UTC All use subject to <http://about.jstor.org/terms>

398 G. DWORKIN AND D. BLUMENFELD:

i.e. reform and deterrence. The question for the act-utilitarian is whether in Case 1 there is any utility in the act of punishing Y more than X. Surely X needs reform just as much as Y and it is hardly plausible to suppose that a harsher sentence for Y will encourage others to commit attempted murder rather than murder (which is an absurd distinction). There seems to be no utility which justifies the disparity in punishment. A general objection that may be raised by any ethical theorist is that the failure of an attempt creates a prima facie case that the intention was altered or withdrawn. This objection depends on an at least questionable psychological theory. But even granted the truth of this claim, it would only be an argument for withholding punishment until it is established whether the failure of the attempt was actually due to the alteration or withdrawal of the individual's intention. After all, we are discussing a case in which, by hypothesis, the failure is due to an accident and not to the alteration or withdrawal of an intention, which is perfectly legitimate unless someone were to support the stronger, but much less plausible view that there are never accidents in cases such as these. It might be thought that the rule-utilitarian could produce a better account of our datum. The question for him is whether there are beneficial consequences of having a rule which specifies heavier punishment for murder than for attempted murder. But what could these benefits be? On this issue the rule-utilitarian is in no better position than the act-utilitarian. With respect to deterrence they are committed to equality of punishment and with respect to reform any inequality would be justified only by the specific psychic make-up of the individuals concerned-responsiveness to

punishment, guilt-feelings-and the difference with respect to punishment would have no connection with the nature of the crime itself. We see here a conflict between the theory and our current practices. We do not suggest that therefore the theories in question must be inadequate. Many will take the existence of this conflict to indicate the inadequacy of the theories, since they feel that the degree of punishment must have something to do with the nature of the act, and thus that punishment must always be more severe for the murderer than for the attempted murderer. But there are those who will conclude that our practices are wrong and will urge reforms equalizing punishment. In the past utilitarians have not done this, perhaps because it is an unrecognized consequence of their position or perhaps because of the feeling that such reforms would be unjust. In the latter case the datum will have proved too obdurate for the theory.

This content downloaded from 131.96.12.74 on Tue, 24 Apr 2018 20:37:41 UTC All use subject to <http://about.jstor.org/terms>

PUNISHMENT FOR INTENTIONS 399

What will the retributivist say about this matter? It is open to him to defend our moral practices by holding that there is a duty to punish in proportion to the amount of harm done. Mr. Y committed murder thus causing more harm than Mr. X. Therefore it is only just to punish Y more than X. This theory fits the datum very closely. But the fit is so tight that the theory, in effect, simply recapitulates the datum. It does nothing to make the datum more intelligible, to provide us with a satisfactory mode of comprehending it. It repeats how we feel about the matter but does not tell us why. One might reply that not everything can be justified, that eventually we reach brute data. There is something in this point. It does seem 'obvious', 'brute', that, e.g. what we do accidentally, not negligently, must be viewed differently from what is done with full knowledge, deliberately. How could one even begin to justify this distinction? Yet even here we can begin to talk about how the person is involved in his acts to a greater extent when they are deliberate, how what is done inadvertently is less connected with the person as moral agent, how the very notion of an action is connected conceptually with the ways in which it can fail, in short, how the moral distinction is connected with the rest of our world. This is not justification in the sense that 'X is right, because . . .' but it does broaden and consolidate our understanding of what is moral. This account is all the more needed here since the admitted difference in harm done between murder and attempted murder had nothing to do with the agent. The amount of harm in the universe cannot be the sole factor involved or else we should punish accidental deeds equally with intentional ones. Justifying a difference in degree of punishment would seem necessarily to involve some reference to a difference in the moral worth of the individuals in question. But, *ex hypothesi*, in our case there is none.

III

How will an act-utilitarian treat Case 2? The question is: Why is Y punished while X is not? If the object of punishment is reform and deterrence X certainly requires reform as much as Y or if this is impossible he must be deterred from translating his intentions into actions at some later time. (Granted in our specific case the latter is not at issue since the President is dead, but

perhaps X is an anarchist and the next President is already on his list.) Also, we have to deter others like X. One possible reply

This content downloaded from 131.96.12.74 on Tue, 24 Apr 2018 20:37:41 UTC All use subject to <http://about.jstor.org/terms>

400 G. DWORKIN AND D. BLUMENFELD:

is that there are other harmful consequences of punishing X which outweigh the beneficial ones. What are these? One might raise a question whether such an act would set a precedent having harmful consequences. In so far as we are considering the precedent set by this particular act of punishment, the arguments raised by an act-utilitarian are identical with those raised by a rule-utilitarian, i.e. ones concerning the consequences of having such and such a rule or general policy. Hence, we shall take up this kind of objection in treating the rule-utilitarian. Another reply, familiar from the recent literature, is that in the absence of any crime committed by X, inflicting pain on X will not count as punishment. But in the first place if a crime is demanded there is one, namely, 'intending to commit murder'. If one objects that this is not a crime, he simply begs the question, since, in effect, we are discussing why it should not be considered as such. The rule-utilitarian can raise a number of objections to having a policy or rule stating that people are to be punished when the only factor preventing them from committing a crime is lack of opportunity. He can suggest that it is very difficult to know an individual's intentions and thus there would be mis-applications of the policy, undermining public confidence in the law and its agents. This is a plausible objection but can be undercut by including in the rule the proviso that only when reliable information about the individual's intentions is available can the individual be punished. And surely there are such cases. Señor X may have written extensively in his diary of his plans, confided in his friends, kept a map of the President's procession, etc. Another objection is that punishment for mere intentions defeats the very purpose of punishment which is to provide the individual with an incentive for not translating his intentions into actions. Punishing intentions would lead individuals to accept the might-as-well-be-hanged-for-a-sheep-as-a-lamb principle. But how is the objector to state his point? Is it supposed that individuals will reason as follows? "If I'm to be punished for intending as well as for doing, I might as well get the satisfaction of doing." This makes no sense. For there is no such thing as mere intention. To have on intention to do X, as opposed, say, to having a fantasy of doing X, IS to be prepared to take steps to do X given the chance. One cannot sensibly say, "Now I'll not confine myself to merely intending to do X given the opportunity, I'll actually attempt to do X when I get the chance." Thus, failing to punish for intentions could not (logically) function as an incentive for individuals to merely intend

This content downloaded from 131.96.12.74 on Tue, 24 Apr 2018 20:37:41 UTC All use subject to <http://about.jstor.org/terms>

PUNISHMENT FOR INTENTIONS 401

such and such but not intend to do such and such given the chance. The objection under consideration depends on such a distinction, and there is none. But there are other and still stronger objections. What would a system of laws embodying a rule providing for the

punishment of intentions look like? When would punishment be administered? As soon as we find out the agent's intentions? But how do we know he will not change his mind? Furthermore, isn't the series-fantasying, wishing, desiring, wanting, intending-a continuum, making it a rather hazy matter to know just when a person is intending rather than wishing? This last objection has two aspects, the difficulty of the authorities distinguishing between fantasying, wishing, etc. and even more importantly the difficulties the individual would have in identifying the nature of his emotional and mental set. Would we not be constantly worried about the nature of our mental life? Am I only wishing my mother-in-law were dead? Perhaps I have gone further. The resultant guilt would tend to impoverish and stultify the emotional life. It should be clear that the act-utilitarian can also raise the above considerations. As we understand these two ethical theories, the act-utilitarian can make use of any considerations that the rule-utilitarian raises, but not conversely. For example, the act-utilitarian can always consider whether a given action will lead to a harmful precedent. In considering the consequences of setting a precedent, he is (like the rule-utilitarian) estimating the value of having a certain rule. Similarly, he can always consider the effect of a given action on generally useful policies, e.g. breaking this promise on promise-keeping in general. The rule-utilitarian, on the other hand, cannot raise the latter kind of consideration. He cannot justify the breaking of a given promise on grounds of utility, although he can consider the question of the justification of the institution of promising on grounds of utility. In the process of justifying a given action, then, an act-utilitarian can appeal to anything a rule-utilitarian can-and more. These are cogent and reasonable objections. The utilitarians seem to be in a good position to justify our current practices as regards Case 2. However, retributivists object to the utilitarian account on the grounds that it simply gives the wrong reasons, i.e. non-moral ones. The reasons offered for failing to punish for intentions are relevant only from 'a practical point of view'. For the retributivist the morally relevant factor is that Y caused harm that X did not. Hence, only Y deserves

This content downloaded from 131.96.12.74 on Tue, 24 Apr 2018 20:37:41 UTC All use subject to <http://about.jstor.org/terms>

402 G. DWORKIN AND D. BLUMENFELD:

punishment. We do not wish to imply that retributivists assign punishment only on the basis of a difference in harm done, ignoring intentions altogether. For most deontologists the harm done must be done with bad intentions if the act is to be morally wrong and thus be punishable. But a necessary (not a sufficient) condition of differentiating individuals with respect to punishment where both have equally bad intentions is a difference in harm done. Our dissatisfaction with this account has already been discussed with respect to Case 1. But there is an insight in the charge that the reasons given by the utilitarian are not moral ones. Unfortunately the difficulty of this objection and others of a similar nature raised against utilitarians is that the distinction 'practical or expedient versus moral' is itself a problematical one. Why are justifications referring to the anxiety, fear, and pain of human beings which are used by the utilitarian to explain Case 2 simply matters of expediency? They certainly do not fit the typical cases falling into this category such as self-regarding actions undertaken without reference to the needs and rights of others. If the sheriff in a small town, dependent for his office on the mayor, ignores evidence that

points to the implication of the mayor's son in a hit-and-run case we have a clear case of acting for expediency. But suppose that another official, in the same position, does not yield to the obvious threat to his office, and investigates the case as he would that of any other citizen. Suppose he acts on the principle of equal treatment of all citizens before the law. This is clearly a moral reason for his position. Now consider a more complicated case. Although he considers the use of narcotics a serious crime, the law-enforcement official in a small mountain community ignores the fact that the town's only doctor is a user. He justifies this action to himself by arguing that if he prosecutes, the doctor will be sent to the State Institution leaving the town without any professional medical care for an indefinite period of time. Was this consideration a matter of morality or of expediency or practicality? The case is similar to that of the corrupt sheriff by virtue of the fact that someone who was guilty of a crime received special treatment and was not punished. But it differs from that case in that the reason supporting the action was not a matter of personal interest but rather the welfare of the community. Whereas no one would call the former a moral reason for withholding punishment, the latter is at the minimum a good candidate. Similar considerations are taken into account in framing the law-when we decide not to make certain acts which are morally

This content downloaded from 131.96.12.74 on Tue, 24 Apr 2018 20:37:41 UTC All use subject to <http://about.jstor.org/terms>

PUNISHMENT FOR INTENTIONS 403

offensive to the community punishable because their nature dictates that the method of investigating them would violate certain other values the community hold dear, such as privacy of the home. To make the matter clearer let us distinguish between two kinds of reasons that may be offered concerning the justification of punishing a given individual. The first tell us something about the individual who is to be punished-"He did it with malice aforethought", "He was old enough to know better", "He'll do it again", "He fully intended to kill him". These reasons reflect the moral worth of an individual. They are tied to his intentions, motives and dispositions. Hence they tell us something about the nature of the person in question. They characterize him. We shall call reasons of this type agent-directed reasons. But, as we have already seen, agent-directed reasons are not the only ones that are taken to be relevant to the question of punishment, e.g. "The community needs him", "To apprehend such people would require 'bugging' their homes, invading privacy". These reasons concern the general welfare, not the moral worth of the individual. They tell us what the consequences of punishing the individual would be to others. Let us call these other-directed reasons. It is crucial for understanding the concept of punishment that we see that these reasons are not on a par. The infliction of pain on a person will not count as punishment unless agent-directed reasons are appealed to in the justification. These reasons are necessary to any justification. Other-directed reasons are neither necessary nor sufficient for the justification of punishment although they can be sufficient conditions for withholding punishment. It is only after punishment is justified on agent directed grounds that we are entitled to raise the issue of the effect of the punishment on society. One has the feeling that in the ideal case, God's judgement, for example, only agent-directed reasons need to be taken into account. Punishment would be based on moral worth. To see the necessity of

agent-directed reasons in the justification of punishment, consider cases in which we attempt to justify " punishment " by appealing only to other-directed reasons. Suppose the general welfare of the community required the sterilization of mental defectives. Whatever the pros and cons of the issue, what we do to these people cannot be called punishment. To this it might be objected that it is not a matter of what we call punishment, but rather a question of what course of action

This content downloaded from 131.96.12.74 on Tue, 24 Apr 2018 20:37:41 UTC All use subject to <http://about.jstor.org/terms>

404 DWORKIN AND BLUMENFELD: PUNISHMENT FOR INTENTIONS

would be right here. Perhaps what we do to these people would not count as punishment. Very well: Let us call it telishment and ask whether telishment is justified.' But this kind of objection only arises when someone asserts that one of the logical conditions for calling something punishment has not been met, to which the reply is " All right, we cannot punish him, we'll telish him ". This issue does not arise for us since all we are arguing is that there are certain intentions which are not punished although they do seem to meet all the conditions for punishment. It is this discrepancy that we are trying to account for.

IV

By drawing the distinction between agent-directed and other-directed reasons we can grasp what leads the retributivist to insist that the utilitarian account is inadequate. The justification that the utilitarian gives is in terms of other-directed reasons. The retributivist sees that these reasons are not the fundamental ones but overstates his point by holding that other-directed reasons are not morally relevant at all. The utilitarian oversimplifies by not explicitly recognizing the two types of reasons, thus making it seem that there is no question of logical priority. By putting the two types of reasons on the same level he leaves himself open to the charge that he could justify punishment solely on the basis of other-directed reasons. Thus, to conclude, in Case 1 we have seen that no ethical theory discussed is capable of giving a satisfactory account of why we always punish attempted murder less severely than murder. In particular no theory gives an adequate explanation of cases in which accident accounts for the failure of an attempted crime. Thus, either the theories are unsatisfactory or our practices should be changed. In case 2, in which lack of opportunity is in question, only the utilitarians have an account of why we do not punish on the basis of moral worth. However, if this account is to be fully satisfactory and if they are to accommodate themselves to the kind of objections made by retributivists, we have argued that the utilitarians must draw a distinction between the kinds of moral reasons to which we have called attention.

University of California, Berkeley

¹ This coinage is borrowed from John Rawls' " Two Concepts of Rules ", *The Philosophical Review* (Lxiv) 1955.

This content downloaded from 131.96.12.74 on Tue, 24 Apr 2018 20:37:41 UTC