

David O. Brink; UCSD  
PHIL 168; Winter 2008  
Handout #6: Punishment

### **PUNISHMENT AND ITS RATIONALE**

If we are to justify or reform our practices of punishment, we must have some provisional idea of what punishment is. Our focus is on **legal** punishment. One view is that a person is legally punished when

- (a) she is deprived of some normal rights or privileges of citizenship,
- (b) by the state,
- (c) for (alleged) infractions of legal rules, and
- (d) after a trial, conducted in accordance with due process.

This appears to tell us **whom** to punish. But it doesn't yet tell us **why** we should punish, **how** we should punish, or **how much** we should punish. A satisfactory theory of punishment should provide an answer to all these questions. How we answer one question may affect how we can answer other questions.

### **REHABILITATION**

Some view crime as the expression of social disease or dysfunction, for which the criminal may or may not be responsible. Either way, this view seems to imply that rehabilitation is or ought to be the main goal of punishment. On this view, the state functions as a kind of therapist whose goal is equip criminals with the sort of personal and social skills necessary to become well adapted and productive members of the community.

If rehabilitation is the exclusive or main goal of punishment, then it looks as if our penal practices are often unjustified. Over-crowded and brutal prisons in which insufficient resources are devoted to education and job training are schools for social pathology, not schools for the social sentiments. No doubt there is need for penal reform, but a huge mismatch between penal rationale and penal practice may give us reason to rethink the adequacy of the rationale.

Moreover, we may doubt whether rehabilitation could be the sole rationale for punishment. Is rehabilitation a good guide as to whom we should punish? Many people, including those who have not broken the law, may be in need of social adjustment. Is the state permitted to require compulsory therapy for those who have not committed crimes? If so, rehabilitation would be an over-inclusive theory about whom to punish. Does rehabilitation explain how we should punish? Should punishment (for the guilty) be restricted to therapy, or is it appropriate to demand some forms of restitution, community service, or suffering for the suffering that the criminal has caused to innocent victims? Also, is rehabilitation a good guide to how long to punish? Suppose we have two people who have committed equally serious crimes for which they are equally responsible. Should their sentences differ just because one is easier to resocialize than the other? And what about those who cannot be rehabilitated? Do they deserve no punishment at all?

## CONSEQUENTIALIST THEORIES OF PUNISHMENT

Rehabilitation is a **forward-looking** rationale for punishment. Another forward-looking rationale is consequentialism, which justifies punishment by appeal to its good consequences, such as its tendency to reduce crime and so to promote peace and security.

Such views about punishment are sometimes part of general consequentialist moral views about what individuals and governments can and should do.

**Consequentialism** (act consequentialism) says that one ought to perform an act iff the value of its consequences is at least as good as any alternative available to the agent (i.e. iff it maximizes value).

As such, consequentialism does not tell us what things are valuable; so different versions of consequentialism result from different conceptions of value. A historically influential version of consequentialism is **utilitarianism**, which understands value as human happiness. Of course, different versions of utilitarianism result from different conceptions of happiness. One influential conception of happiness is **hedonism**, which claims that pleasure is the only intrinsic human good and pain is the only intrinsic human evil.

Jeremy Bentham (1748-1832) is a hedonistic utilitarian, and he offers a utilitarian rationale for punishment. Punishment has hedonic costs, not only for the person punished, but often for his associates (many of whom are innocent), and the institutions of corrective justice -- the police, court, and prison systems -- are expensive, unsafe, and sometimes unreliable. So, as Bentham insists, punishment must be justified. He appeals to its expected benefits, especially to its tendency to reduce crime. Punishment can have two kinds of deterrent value.

- **General Deterrence.** Punishing A for his crime has value as a general deterrent insofar as doing so tends to deter others from committing similar crimes.
- **Specific Deterrence.** Punishing A for his crime has value as a specific deterrent insofar as it prevents A from repeat offense.

Other ways in which punishment might reduce crime include rehabilitation and moral education. Punishment might also have benefits other than those of reducing crime, such as the satisfaction that victims and others might experience at seeing the guilty suffer. However, deterrence seems to be the good on which consequentialists often focus; indeed, deterrence would be the view that this is the only relevant consequence for justifying punishment.

A consequentialist of Bentham's stripe also has a natural answer to the question about how and how much to punish.

- We should punish in ways calculated to deter crime, and the severity of punishment should be such that it is greater than the expected profit of each offense discounted (divided) by the perceived probability that the infraction will be punished.

## IS A PURE CONSEQUENTIALIST THEORY OF PUNISHMENT OVER-INCLUSIVE?

Whereas consequentialists justify punishment in **forward-looking** terms, retributivists justify it in **backward-looking** terms -- someone should be punished iff he deserves punishment for past crimes. Retributivists claim that consequentialism leaves out this crucial backward-looking dimension. In particular, they worry that consequentialism might sanction punishing the

innocent. The stock example is that in response to a recent crime spree a sheriff may be able to prevent a security crisis and general unrest if he frames an innocent person for the crimes. The anti-consequentialist notices a tension in the following triad.

1. The state ought always to maximize value/utility.
2. It is never permissible for the state to punish an innocent person.
3. Sometimes it would maximize value/utility to punish an innocent person.

Something has to go. The anti-consequentialist invites us to reject (1).

One could question (3). The example appeals to considerations of general utility, including peace of mind. But this appeal seems imperfect. It does not factor in the unrest that would result if the deceit is discovered. And the deceit is likely to be discovered, because even if it promotes general deterrence, it clearly does not promote specific deterrence (the actual criminal is not punished). But then the crime spree is likely to continue, and when it does the deceit will be exposed. But then the original anxiety about crime will return and now will be accompanied by anxiety about the possibility of being framed by the local police.

Moreover, the consequentialist may not appeal to general utility but may appeal to the special value of punishing wrongdoers or protecting the innocent and claim that neither value is promoted by framing the innocent.

But the anti-consequentialist may think that this misses the point, because it might sometimes happen that if the state has to punish an innocent person in order to minimize punishment of the innocent. Suppose that Jethro McCoy fell on his own knife but that the only witness to this is Buford Hatfield, whom no McCoy will believe did not kill Jethro. Suppose that feuding code requires that killings be avenged fivefold when left unpunished by the state. If so, to avoid the McCoy's killing five innocent Hatfields (including Buford), it seems necessary to frame an innocent Hatfield (e.g. Buford). Maximizing protection of the innocent seems to require harming (punishing) the innocent.

I suppose one could question this new case for (3). But the anti-consequentialist is claiming that whatever the consequentialist values, there seem to be imaginable circumstances in which punishing the innocent would be best overall. Moreover, the anti-consequentialist assumes that even in such circumstances we will hold onto (2). We could question this. One might support (2) by appeal to the value of protecting the innocent. But bad as it is to punish an innocent person, it's worse for five innocent people to suffer comparable harm. If we care so much about protecting the innocent, isn't it irrational not to minimize their suffering, which sometimes may require causing suffering to some innocents to protect more? It's not obvious what to say here, which is perhaps reason enough to explore other theories of punishment that preserve (2).

### **RAWLS' RULE CONSEQUENTIALISM**

In "Two Concepts of Rules" (1955) John Rawls modifies the utilitarian theory of punishment in a way that he thinks avoids this retributivist worry. Rawls asks us to distinguish the **legislative** issues of whether to punish conduct, which conduct to punish, and how to punish such conduct from the **judicial** issue about the conditions under which particular individuals ought to be punished. This, he thinks, is a special case of the more general distinction between the reasons for having a practice and the reasons for following its rules. He thinks that forward-

looking utilitarian considerations are relevant to addressing these legislative issues, but that only backward-looking retributive considerations are relevant to addressing the judicial issue.

The decision whether or not to use law rather than some other mechanism of social control, and the decision as to what laws to have and what penalties to assign, may be settled by utilitarian arguments; but if one decides to have laws then one has decided on something whose working in particular cases is retributive in form [684].

This looks like a **mixed** theory of punishment, but it fits with one traditional way of formulating utilitarianism. Act utilitarianism is a form of **direct** utilitarianism, because it assesses everything it does (actions, persons, policies, etc) by the value of its consequences, when compared with alternatives. Whereas act utilitarianism tells us to do the best action, rule utilitarianism tells us to act in accordance with the best rules. This makes rule utilitarianism a form of **indirect** utilitarianism. Whereas act utilitarianism applies the test of consequences directly to acts, rule utilitarianism applies the test to rules and then instructs you to act on the best rules. Rawls thinks that rules allowing the framing of innocent people would clearly have worse consequences overall than rules permitting the punishment of only those who are guilty.

There are at least two worries about Rawls's proposal, which are general worries about rule utilitarianism. First, one might wonder whether one really can separate the legislative and judicial issues so neatly. The legislative issue is what rules to have, and we might wonder why utilitarians shouldn't prefer rules that allow punishing the innocent when this is necessary to minimize harm to the innocent. Also, one might wonder why a consequentialist should be a rule consequentialist. If rule consequentialism and act consequentialism are different, then it must be true that by acting on the best rules I sometimes perform suboptimal acts. But if consequences are what matter, why not apply the utilitarian principle directly to acts and always perform the best act?

## RETRIBUTIVE THEORIES OF PUNISHMENT

Perhaps the pure consequentialist account is defensible, but it is troublesome enough to justify exploring retributive theories of punishment in their own right. Retributive theories answer the questions about who should be punished and why by insisting that punishment be reserved for those who deserve sanctions on the basis of prior **wrongdoing**. They answer the question of how and how much to punish by appeal to the idea of "an eye for an eye" or some notion of **proportionality**. For example, Immanuel Kant says that the utilitarian willingness to punish the innocent treats that person as a mere means and so violates the fundamental norm of respect for other people. Respect for others, Kant claims, requires holding them responsible for their free choices. This requires that we punish all and only the guilty. Kant also thinks that a norm of equality requires that punishment be proportional to the harm caused.

We might notice some worries about these retributive ideas. First, to understand how the guilty are morally deserving of punishment, we must apparently understand their guilt as moral guilt or wrongdoing. But not all moral guilt seems ripe for punishment. Perhaps God should or will punish all moral wrongdoing, but it doesn't seem appropriate for the state to punish wrongdoing that has not been criminalized. Indeed, Kant suggests that "inner wickedness" ought to be the basis of punishment. But this would apparently go even further, allowing punishment for bad character and not even requiring bad acts. Second, we might note problems with some interpretations of the proportionality doctrine. Taken literally, the principle of an eye for an eye

is quite implausible. It would require us to steal from those who steal, to break the arms of those who break arms, and to rape rapists. In the face of such worries, retributivists often retreat to a proportionality formula. For example, Robert Nozick claims that

The punishment deserved depends upon the magnitude  $H$  of the wrongness of the act, and the person's degree of responsibility  $r$  for the act [on a spectrum 0-1, in which 0 indicates no responsibility and 1 indicates 100% responsibility], and is equal in magnitude to their product,  $r \times H$  [Philosophical Explanations, p. 363].

Nozick seems to think of the magnitude of the wrong as proportional to the harm that it causes its victim. But many crimes cause no harm at all. For instance, attempted murder and failure to report a bribe, though crimes, need cause no harm. Or suppose that A breaks the hands of a garbage collector whereas B breaks the hands of a concert pianist. A causes more harm, but should he be punished more severely? Perhaps we could understand Nozick's proposal as making Punishment proportional to Desert, which is itself the product of Wrongdoing and Responsibility.

$P$  is proportional to  $D = W \times R$

### **PUNISHMENT, FAIRNESS, AND CONTRACT**

We can avoid some problems for retributivism if we punish for legal, rather than moral, guilt. But why should we do this? It seems easy to say why people are morally deserving of punishment for moral wrongdoing. But why are they morally deserving of punishment for legal wrongdoing? In "Persons and Punishment" Herbert Morris sketches an answer that appeals both to notions of **fairness** and to notions of **contract**. In doing so, he sees himself as developing the Kantian appeal to respect for the criminal's agency. Indeed, Morris says that he is defending the somewhat paradoxical (Hegelian) claim that the criminal has **a right to be punished**.

One line of argument appeals to respecting the criminal as a responsible agent. We treat some as responsible when we permit him to make informed choices that determine his fate. Treating criminals as agents involves subjecting them to the sanctions that were advertised as the consequence for failure to comply with the rules of the criminal law. On this view, punishment is just giving the criminal what he freely chose to risk. Morris contrasts this sort of recognition of the criminal's agency with the view of the criminal as a patient or object of therapy implicit in the rehabilitative view of punishment.

Morris also appeals to the unfairness that the criminal practices and that demands redress. Morris's idea is that a reasonably just criminal code establishes norms of forbearance -- for instance, against theft, assault, and murder -- the general observance of which is mutually advantageous. We each benefit from other people's compliance with such norms, and fairness requires that we do our part to maintain this system of restraint. The criminal profits from others' compliance without shouldering his own share of the cost of the system. Criminal behavior represents taking more than one's fair share. Equality of shares requires depriving the criminal of these advantages. In principle, one could do this by the criminal giving up his advantage or, where this is impossible, providing restitution. Perhaps this is where the contractual argument is supposed to kick in. The criminal code not only specifies mutually beneficial norms of forbearance but also specifies penalties for noncompliance. Treating criminals as responsible

agents requires responding to their taking more than their fair share by enforcing the terms for enjoying the benefits of others' compliance.

One question is whether a contractual approach to punishment requires any upper limit on punishment provided that punishment is announced in advance. Intuitively, it seems disproportionate punishment to feed pickpockets to the sharks. But if punishment is just what the criminal contracts for, how can there be anything wrong with this punishment provided the criminal is aware of it in advance and has a fair opportunity to comply? Suppose I throw my wallet into shark infested waters and Billy Bob jumps in after it and is eaten by the sharks. It's not clear that I've done anything impermissible (though you might not want me to babysit your children). After all, I didn't make Billy Bob dive in; he chose to. But how is this different from the state setting the penalty for pick pockets at walking the plank over shark infested waters? If the contractual theory is right, can the criminal really complain about disproportionate punishment? But if the contractual form of retributivism cannot explain why there is an upper limit on justified or proportional punishment, then this may be a worry about that form of retributivism.