

**Phil 108, August 10, 2010**  
**Punishment**

**Retributivism and Utilitarianism**

**The retributive theory:**

- (1) It is good in itself that those who have acted wrongly should suffer. When this happens, people get what they *deserve* and *justice* is done.
- (2) A justification of punishment is that it realizes this good, by inflicting suffering on those who have acted wrongly (ie., by giving criminals what they deserve, by letting justice be done). Stronger versions of this thesis are:
  - Realizing this good should not be sacrificed for any other social aim. (“Let justice be done though the heavens fall!”) For example, we should not impose *more lenient* punishments that might be *more effective at reforming* criminals.
  - Realizing this good is the *only* justification of punishment.

**Bentham’s utilitarian theory of punishment:**

- The retributive theory is false.
- What is good in itself is just that there is more pleasure and less pain: that the sum total of pleasure less pain, taking everyone into account, is higher.
- Social policy should aim to realize this good: to maximize the sum of pleasure less pain.
- In itself, punishment is bad, period. The person punished feels pain, and his pain counts just as much as anyone else’s in the sum total.
- Punishment can only be instrumentally good: good because it has good effects. For Bentham, these good effects are more pleasure and less pain.
- Punishment may have the good effect of *preventing* acts that would produce *pain*.
- Punishment prevents such acts by leading other people to expect that if they act in the same way, they will suffer pain just like the person punished. This expectation *deters* them from performing such acts.
- Punishment is justified only if its good effects—principally, the pain that is avoided by deterring people from performing actions that would cause it—outweigh its bad effects—principally, the pain visited on the punished person.

**Attractions of the utilitarian theory of punishment:**

- (1) Humane. Punishment at best a necessary evil.
- (2) Seems to explain many accepted exceptions and qualifications. According to Bentham, we should not punish when punishment is:
  - (i) *Groundless*: where the act to be punished does not produce pain or foreclose pleasure. E.g., consensual acts.
  - (ii) *Inefficacious*: where the punishment will not prevent the act. E.g., *ex post facto* or unannounced laws; children and the insane; unintentional actions, involuntary movements, duress.
  - (iii) *Unprofitable*: where the punishment has worse effects than the acts it aims to prevent.
  - (iv) *Needless*: where the punishment has worse effects than some other means of preventing the same acts.

- How might the retributivist try to explain these exceptions and qualifications? One example: The insane should not be punished, because they do not deserve to be punished, because they are not responsible for what they do, because they do not act freely and intentionally. Notice while both the retributivist and the utilitarian appeal to the fact that the insane “know not what they do” to explain why they should not be punished, this fact plays altogether different roles in their respective explanations.
- (3) Seems to explain the “proportionality” of punishment: Why we should not impose, e.g., the death penalty for jaywalking. And it gives us a clear way to determine what the proportion is.
- *Exercise*: What is this way? How does utilitarianism determine what the proportion is?
  - *Question*: The retributivist may be able to explain proportionality. But how can the retributivist determine what the proportion is? (See Ewing, p. 36ff.)
- (4) Avoids metaphysical worries: If *determinism*<sup>1</sup> is true, then we are we ever *really* responsible for what we do?<sup>2</sup> And if we are never really responsible for what we do, can we ever deserve anything? If punishment is justified only to give people what they deserve, and if people never deserve anything, then punishment is never justified. By contrast, if punishment is justified to deter people from bad actions, then punishment can be justified even if no one deserves anything.

#### **A problem with the utilitarian theory of punishment:**

*Example*: Framing an innocent man to prevent a riot. In general: “Scapegoating,” “making an example of,” etc., might lead to greater pleasure and less pain.

How, on the utilitarian theory, does scapegoating differ from punishing a guilty person to deter others? In both cases, we are inflicting pain on one person in order to save others from greater pain, and according to the utilitarian, these facts about pleasure and pain are the *only* morally relevant facts. Fact about innocence, guilt, desert, justice, etc. are irrelevant.

The retributive theory does not have *this* problem.

- There is *no* reason *to* punish an innocent man, since doing so does nothing to achieve justice: to give bad people the suffering they deserve.
- Moreover, if *good* people deserve *not* to suffer, then there is additional reason, besides the pain inflicted, *not* to punish the innocent man: it would be *unjust*.
- However, this last point also poses a *problem* for the retributivist. When we punish the guilty, don’t we unjustly inflict suffering on his innocent family? (Ewing, p. 43)

---

<sup>1</sup> The thesis that all events are causally determined by prior events. This means that our actions were ultimately causally determined by events that occurred before we were born and so over which we had no control.

<sup>2</sup> See Ewing, pp. 35–6 for some suggestions on how the retributivist might just answer “yes” to this question.

## **Quinn: The right to punish from the right to threaten**

### **The basic question:**

Why doesn't punishment violate (or even infringe) the rights of the person punished? After all, taking someone's property, confining him against his will, etc. normally does violate his rights (or infringe them, in which case we owe compensation or apology).

The same question arises in the case of self-defense, but it seems easier to answer: we defend ourselves to protect our rights. The problem is that when we punish someone for some violation of our rights, the *punishment* does not protect us from *that* violation.

*Objection to deterrence:* It is not generally true that "ordinary rights to liberty and life fail to apply *because* their application would stand in the way of some socially profitable use of people."

*Objection to forfeiture:* Forfeiture is the idea that rights to life, liberty, and property are conditional, so that if one attempts to violate the rights of others, one "forfeits" them. But we don't think that people or society can violate the criminal's rights in ways *independent* of punishing him (e.g., locking up a burglar to prevent him from expressing his political views). So "the idea of forfeiture... comes to no more than the idea that the criminal's rights do not in fact stand in the way of his being punished."

*Objection to retribution:* It is mysterious why it is just to inflict suffering for wrongdoing. And the main appeal of the view—that punishment should be justified by the crime, not by further advantages—can be captured by a new alternative.

### **Quinn's approach:**

*The traditional approach:* The right to threaten punishment is justified by the right to punish.

*Quinn's approach:* The right to punish is justified by the right to threaten punishment.

*To create a threat:* (i) to create a risk that something bad will happen to someone if he does something and (ii) to inform him of (i) in circumstances in which this information may deter him from doing it.

*Self-protection:* We have a right to create threats because we have a right to protect ourselves from violations of our rights.

- Like deterrence theory, punishment is justified by the aim of deterrence. But each act of punishment is not justified *because it* deters.
- Like retribution, the justification of punishment is backward-looking, for a particular crime. But the justification does not appeal to desert.

Our right to protect ourselves from violations of our rights allow us to take several kinds of measures:

- (i) self-defense,
- (ii) barriers (e.g., locks, fences),
- (iii) automatic costs that "*precede or accompany* the violation of some right,"

- (iv) confinement, as a last resort.

Creating threats is an instance of (iii): “[S]uppose... our defender cannot arrange the spikes so that they offer a threat of injury to someone entering his territory but can arrange them so that they clearly offer threat of injury to an enemy leaving his territory after an attack.” If he would have a right to do the former, doesn’t he have a right to do the latter?

### “M-punishment”

With a system of mechanical punishments or m-punishments, our only choice is whether to create the threat. There is no further choice, after we have created the threat, whether to follow through. The machines take care of that for us.

Each instance of m-punishment is justified *not* because *that punishment* deters *future* crime. Instead, it is justified as the *unavoidable by-product* of the prior threat, which was justified because *that threat* deterred crime *until now*.

*Proportionality* in m-punishment is explained by proportionality in self-protection.

- For example, we are not allowed to kill someone trying to pick our pocket.
- Still, the cost we threaten may be more serious than the violation of the right that the threat protects.

### The functional equivalence of punishment and m-punishment

“Every intuitively justified practice of punishment has as its counterpart a practice of m-punishment justified by the rights of self-protection, and vice versa,” where they are “counterparts” if they “threaten penalties of just the same severity for the same crimes.”

Since m-punishments may be more severe than the violations that they punish, one might worry that the system of m-punishment is at odds with our intuitions about proportionality.

- *Reply*: Intuitively, we allow that punishment may be more severe than the violations they punish.

There are cases in which a violation of rights has occurred, but punishment seems unjustified. It is also unjustified to m-punish in these cases?

- *Innocent third parties*: M-punishment would also be ruled out. “[N]o matter what the gain in protection, it is manifestly unjust to threaten to inflict an evil on someone when it is not up to him to do that which will prevent it.”
- *Incapacity to respond to reasons* (e.g., compulsion, mental illness, non-culpable ignorance): “It would be unjust to create dangers for them that they cannot escape or cannot have a reasonable chance of escaping.”

### Does the right to threaten punishment justify the right to punish?

Quinn aims to argue from:

- (1) the right to threaten *m*-punishment

to

- (2) the right to *threaten* punishment

to

### (3) the right to *punish*

The hard step is the one from (2) to (3). Can this be explained *without presupposing* the right to punish? Notice that this is the crucial difference between m-punishment and punishment. In the case of m-punishment, there is no question of the right to *inflict* m-punishment, because it is the machines that do that, not us. We only need to show that there is a right to *threaten* m-punishment. But in the case of punishment, *we* inflict the punishment. With what right?

Quinn's explanation of the step from (2) to (3) is rather subtle and involved. So we won't try to follow all the twists and turns. Perhaps the most important feature of his explanation is that it focuses on the *objections of the recipient* of the punishment. From the *recipient's* perspective, there seems little difference between our creating the threat and our trying to carry it out. In both cases, some potential cost is attached to some conduct of his. So if he doesn't have an objection to the former, then why should he have an objection to the latter?

*Question:* Is the *recipient's* objection to our being *allowed* to threaten or to punish the only objection that matters to whether we have a right to threaten or to punish?

- One might think that whether we have a right to do something depends on a balancing of the *recipient's* objection to our being *allowed* to do it to him and *our* objection to our being *prohibited* from doing it to him.
- And perhaps even if *his* objection to our being allowed to punish him is *no stronger* than his objection to our being allowed to threaten him, *our* objection to being prohibited from punishing him is *weaker* than our objection to our objection to being prohibited from threatening punishment. If we don't threaten, then our rights are put at risk. But what do we lose if we don't follow through on our threat? (Of course, we lose the ability to *deter others* by punishing him, but the *right* to punish him is not supposed to depend on that.)

Quinn considers a related objection: that his theory faces a problem similar to the "toxin puzzle."

- In the puzzle, one has reason to intend to drink the toxin, but no reason, when the time comes, to drink it. Once one realizes this, one cannot rationally intend to drink it.
- Why isn't it the same with punishment? We have reason to intend to punish, but no reason, when the time comes, to punish. So we cannot rationally intend to punish, which threatening to punish involves.
- Quinn concedes that our reason, when the time comes, to punish cannot be to deter the crime that we are punishing.
- Still, we can have *other* reasons to punish: to mete out retribution, to deter others, to express our condemnation, or simply to fulfill our contractual obligations.
- Crucially, these reasons do *not justify our right* to punish. But they are reasons that we can expect to have, and so they can enable us to intend in advance to punish.