

**Phil 116, April 26, 2008**  
**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 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.