Session 18 Rawls, "Two Concepts of Rules"

Act-Utilitarianism: An act is wrong if and only if it would fail to produce as much welfare as any alternative act open to the agent.

Rule-Utilitarianism: An act is wrong if an only if it would be forbidden by the set of rules whose (universal? near universal?) adoption would produce the most welfare.

Utilitarianism about Institutions: Our institutions should be designed so that they benefit society, by producing at least as much welfare as any other design would have produced.

- Which version of utilitarianism is Rawls defending? Might the view he defends here best be understood as a kind of *selective rule utilitarianism/consequentialism*?
- Can *rule-utilitarianism* be motivated? (If what we care about is maximizing happiness/good consequences, doesn't sticking to the best rules even when doing so in the particular case will sacrifice happiness look like *rule-worship*?)

Punishment

Defining the Institution of Punishment: Rawls: "a person is said to suffer punishment whenever he is legally deprived of some of the normal rights of a citizen on the ground that he has violated a rule of law, the violation having been established by trial according to the due process of law, provided that the deprivation is carried out by the recognized legal authorities of the state, that the rule of law clearly specifies both the offense and the attached penalty, that the courts construe statutes strictly, and that the statute was on the books prior to the time of the offense." (p. 10)

- Rawls asks whether utilitarian arguments could be used to justify institutions much different from this (and intuitively much more cruel and arbitrary).
- But we might also wonder whether utilitarian arguments could justify individual acts that depart from these norms. This difference, between justifying a practice/institution and justifying a particular act it might be thought to govern, will become central to Rawls' argument later on.

Two theories of just punishment:

Retributivist: Punishment is justified if, only if, and because the person who does wrong *deserves* to be punished. A world in which people suffer in proportion to their wrongs is *morally better* than one in which people don't.

Some features of the *retributivist* view:

- the punishment must be proportional to the crime
- we are never justified in punishing an innocent person
- in evaluating the justifiability of punishment, we should look backwards, to see what the person to be punished did, not forwards, towards the consequences of our punishment

Utilitarian: Punishment is justified if, only if, and because punishing someone will lead to at least as much welfare as not punishing. The severity of the punishment should be determined by what level of punishment would have the best effects.

Some features of the *utilitarian* view:

- because punishment causes (at least temporary) reduction in the welfare of the person being punished, we should always choose the minimal punishment necessary to achieving the good effects
- in evaluating the justifiability of punishment, we should look forwards, towards the consequences of our punishment, not backwards, to the crimes committed

An apparent problem for the utilitarian view:

- Might it justify punishing an innocent person? (See Carritt's example, p. 10)
    - An inconsistent triad?
        (i) The state ought to act so as to maximize welfare.
        (ii) The state ought never to punish a person it knows to be innocent (or release without punishment a person it knows to be guilty?).
        (iii) Sometimes maximizing welfare will require the state to punish a person it knows to be innocent (or release without punishment a person it knows to be guilty).
    - Which of these three propositions can we reject?

Rawls' suggestion: The utilitarian approach to punishment is appropriate when evaluating/justifying the *institution or practice* of punishment (these arguments should influence the *legislator*); the retributivist approach is appropriate when evaluating/justifying a particular decision *within* the practice (these arguments should influence the *judge*). In other words, *utilitarian* arguments justify adopting a *practice of punishment* that is governed by *retributivist rules*.

- *Question*: Would utilitarian considerations lead us to adopt an *institution* of punishment governed by retributivist rules? Would utilitarian considerations definitely lead us to reject "*telishment*"? Would an institution of punishment designed on the basis of utilitarian considerations look any different from the retributivist model? (If they look the same, might they do so only *contingently*?)
    - Will all and only those acts we may have good utilitarian reasons to restrain/deter be those acts we think *wrong* and deserving of punishment?
        - E.g., what if (as a contingent matter) a society reacts with "terror and alarm" to acts which are (or would be, if the society did not react this way) harmless? (E.g., homosexual behavior in a society significantly more homophobic than ours…)
        - What about acts which do spread terror and alarm and are harmful but are not wrong (e.g., being the carrier of a contagious disease)?
    - There may be ways in which the "proportionality" recommended by forward-looking considerations of utility doesn't match up with the "proportionality" recommended by desert

- We might need harsher punishments to deter high-profit, hard-to-detect crimes, even if they seem less bad/blameworthy (e.g. tax evasion v. murder).
- The benefits of incapacitating repeat offenders might justify long prison terms for, e.g., chronic petty thieves, but a comparatively short term for someone guilty of a one-off, very serious offense (e.g., patricide).

  - But other considerations will tend to make the proportionality recommended by considerations of utility match up with that recommended by considerations of desert: as Bentham notes, the punishment cannot be more costly than the crime (or it would be "too expensive"), and the system of punishment should be designed so as to provide incentives to prefer the less harmful crime to the more harmful.
  - Does the *wrongness* of a crime – and the punishment it deserves – always correlate with how *harmful* it is, or might be expected to be?
  - Retribution may be built into the *concept* of punishment. If so, perhaps we can't ask whether it would make sense (from the utilitarian perspective) to have a sytem of punishment that allows for punishing those who aren't guilty (but doesn't our actual system do this?). But we can ask whether a utilitarian would necessarily prefer a system of *punishment* to some other method of crime-control.

- *Question*: Would it be possible/appropriate to justify the *institution* of punishment on purely retributivist grounds?
  - The proportionality rule *underdetermines* the severity of punishment, and "and eye for an eye" seems a bit barbaric.
  - We don't want to punish all wrong acts (at least not through the legal system).

- Rawls suggests that different roles for the utilitarian and retributivist justifications for punishment can be brought out by considering the difference between two kinds of questions:

(i) Q: Why was J thrown in jail? (What justified throwing J in jail?)
A: He robbed a bank. (Backward-looking, retributivist justification)

(ii) Q: Why throw bank robbers in jail?
A: Doing so deters future crimes, and so is for the good of society. (Forward-looking, utilitarian justification)

Are these really two different kinds of questions, demanding two different kinds of answers? Mightn't we ask the first, even if we knew J robbed a bank?

{Promising

The problem for utilitarians: It seems like a utilitarian will have difficulty accounting for our intuition that we are not justified in breaking a promise just because doing so will have the best effects overall (even taking into account the negative effects on the practice of promising), because the act-utilitarian principle, applied to individual decisions about

when to keep or break a promise, *would* tell us to break it when doing so would have the best effects overall. (Consider the death-bed promise case.)

Rawls' suggestion: utilitarian considerations should be appealed to to justify the *practice* of promising, but the rules of a practice justified on utilitarian grounds *would not include* a rule permitting us to break our promises when doing so would be best on the whole, because such a practice would be much less useful than our actual practice. So utilitarian considerations may *not* be appealed to to justify particular decisions falling under the practice.

Again, compare two questions:

(iii) Q: Why did J do that? (Why should J do that?)
A: He promised he would. (Backward-looking)

(i) Q: Why do what you promised?
A: Because it leads to the best consequences. (Forward-looking)

But again, we might ask, are these really different kinds of questions, demanding different kinds of answers?***}***

Practice Rules v. Summary Rules

Summary Rule: a "rule of thumb" or rule of convenience; it represents a kind of "summary" of past decisions arrived at by direct application of more basic reasoning (e.g., the utilitarian principle)

- useful to us because similar cases tend to reoccur
- decisions made on particular cases are logically prior to the rules: "the performance of the action to which the rule refers does not require the stage-setting of a practice of which this rule is a part"
- in principle, we're "always entitled to reconsider the correctness of a rule and to question whether or not it is proper to follow it in a particular case (*But: can a rule of thumb serve a useful purpose if we can question it every time? Mightn't we think, in the case of summary rules, too, that we can only question whether it's a rule we have reason to adopt, not whether we should use it in this case?*)
- employing a summary rule is justified if and because applying it will lead us to be more likely to make an independently correct decision then we would by direct application of the more basic reasoning; but what the right thing to do is in this case is independent of what the rule tells us… *(Though again, perhaps the usefulness of the rule changes our reasons – gives us reasons to follow it even in the case where it, in a gods-view sense, leads us astray…)*
- Examples? Don't tell a fatally ill person that he's dying; Look both ways before you cross the street…

Practice Rule: Rule setting up offices, or defining certain actions and specifying when they are appropriate, establishing penalties, etc.

- Rules of practice are logically prior to particular cases of actions governed by those rules – e.g. striking out, stealing a base, making a promise, punishing: there is no way to strike out without following the rules of baseball. *(But what about cheating?*

I can't get out of criticism for breaking a rule by saying that, since I broke the rule, I was in fact playing a different game and had no obligations defined by it…)

- Someone who wants to perform an action governed by a practice rule can't meaningfully ask whether or not he should follow the rule in this case (consider "can I have a fourth strike?" – such a person would be most charitably interpreted as asking *what the rules were* – see discussion of baseball on pp. 25-26)
- Questioning whether to follow a rule, Rawls suggests, must take the form of questioning the rule itself – that is, questioning whether the *practice* is designed as well as it might be (see footnote 25, pp. 28-9, for an example)
- Defenses of particular actions falling under practices must take the form of appeals to the rules of the practice, and then defenses of the practice as a whole. (Remember our two questions…) Actions governed by practice rules aren't correct or incorrect independently of the place-setting of the practice. (*But: can't I respond by saying, I know J robbed a bank, and I know there are good justifications for a (general) practice of punishing bank robbers by imprisoning them, but ought we to imprison J? After all, imprisoning J is an action that's not defined by the practice, even if punishment is…)*
- "One can be as radical as one likes but in the case of actions specified by practices the objects of one's radicalism must be the social practices and people's acceptance of them." (p. 32)

Questions:

- Are there important disanalogies between punishment and baseball?
    - The role of the legislator and the role of the judge are less cleanly separated in the case of punishment than the role of the designers of the game and of the umpire in baseball.
        - Consider the interpretive work that some legal theorists at least think is the proper role of judges – might it be part of the proper role of judges, in interpreting the law, to take account of some of the same considerations that should move legislators when they frame the law? (Remember Dworkin's "policy" considerations…)
- Is it *never* appropriate to ask whether a particular act mandated by a practice is justified, if we're agreed that the practice itself is justified?
    - I might not be able to ask whether, say, a fourth strike is justified, but I can ask whether my throwing the kid the ball again is justified…
    - I may have good reason to *step out of* the office defined by the practice…
    - It may make an important difference *why* I want to break the rules: consider the difference between internal and external goals…
        - It might sense to break the rules of baseball in order to have more fun in a particular case (think of my birthday-kid), even if I think baseball would not, in general, be more fun if it allowed 4 strikes in special cases. But it wouldn't make sense to break the rules of baseball in order to "get a hit" – since that (internal) goal is one I can achieve only by sticking to the rules of baseball.
- What is *cheating*? E.g., what differentiates cross-checking from attacking someone with a stick in ice-hockey?

24.235J / 17.021J Philosophy of Law
Spring 2012