

Retributivism and uncertainty

Why do we have a criminal justice system? What could possibly justify the state punishing its citizens? Philosophers, scholars of law, politicians and others have proposed different justifications, one of them being retributivism: the view that we ought to give offenders the suffering that they *deserve* for harming others. However, intentionally harming other people and making them suffer is serious business. If we are to do this in the name of what's right and good, we better be very certain that what we do *really is* right and good. The so-called *Epistemic Argument Against Retributivism* calls this certainty into question.

Different criminal justice theories

In the criminal justice ethics literature, a large number of theories of justification have been proposed. It's impossible to do full justice (pun intended) to this literature in a blog post, but below I list some well-known and influential ideas.

1. Deterrence. When the criminal has paid their fines or gotten out of prison, they won't dare do what they did again from fear of consequences. Other people will look at what happens to criminals and be deterred from committing crimes in the first place.
2. Rehabilitation. Here too, the criminal justice system is considered justified insofar as it manages to bring down crime rates. But instead of focusing on scaring people away from crime, rehabilitative approaches focus on giving prisoners the tools to stay away from crime when they get out again. Inside prison, they have access to therapy, education, real job training, and so on.
3. Expressivism: By punishing people who commit certain acts, the state expresses its judgment that these acts are wrong and must not be done: it sends a message of moral values to its citizens.
4. Retributivism: When people commit bad acts, they *deserve* to have something bad happen to them in return. The criminal justice system's job is to deal out just deserts.
5. Restorative justice. This is sometimes conceived of as a version of criminal justice, sometimes as an alternative to it. The main idea is that crime tears a community apart, and real justice restores community again. Restorative justice proponents often argue for having criminals meet with their victims (in case of economic crime, someone who, e.g., cheated on his taxes might be made to confront the realities of public health care and how they struggle with scant resources). There's an emphasis on first making the criminal ashamed of themselves and how they have let their

community/family/colleagues/others in their life down, and then offer them a path for reintegration, often through community service somehow connected to the crime.

In real life, politicians and others often appeal to mixed motivations. It is, for instance, common to appeal to both retributivism and deterrence among politicians who want to be “tough on crime”. It is also common to have a little bit of retributivist thinking mixed in, even when much stress is placed on, e.g., rehabilitation. I’ve talked to American colleagues fascinated by the Norwegian prison system with its strong rehabilitative efforts, and comfy and homely in-prison environments, but in interviews with Norwegian prison wardens, they still talk of *loss of liberty* as a punishment.

Criminal justice philosophers are probably more likely than others to stick to one single theory of justification. However, although there’s a lot of theoretical disagreement about the proper justification for a criminal justice system, there’s more – far from total, but *more* – agreement to be found on a practical level. Philosophers embracing a variety of theories have argued for a much milder and more humane criminal justice system than what is currently the case in countries such as the USA.

Deterrence theorists are ultimately concerned with having a low crime rate in society. Among politicians and laypeople, it’s common to believe that this is best achieved through really harsh punishments. Philosophers and other scholars are often better at taking actual research into account, which shows the importance of more rehabilitation-directed punishments and creating a less criminogenic society. Expressivists point out that what something expresses is relative to the culture in which it takes place and its norms. When it comes to punishment, a society which deals out many lifelong prison sentences and have capital punishment will express that a certain crime wasn’t that serious if the perpetrator is sentenced to one year in prison. But in a society where few crimes are punished by prison at all, a one-year prison sentence might send a serious message. Therefore, we could very gradually decrease and/or change punishments, and yet send the same messages (at a lower price in terms both of money and human suffering). Some philosophers arguing for retributivism believe that what criminals ultimately deserve is to *feel guilty* about what they have done, and therefore embrace something much like restorative justice, but on retributivist grounds.

For the above reason, I think it’s useful to specify, when arguing for reform in countries with harsh and largely retributivism-motivated criminal justice systems, that one argues against *harsh* retributivism. In my article “Retributivism, justification and credence: The epistemic argument revisited”, I used the following, admittedly somewhat loose, definition:

A state has Harsh Retributivism if the primary justification for punishing offenders is the dealing out of just deserts. Deterrence and other considerations might still play some role, but desert is considered most important. Furthermore, crimes (at least the more serious ones) are punished harshly. Offenders receive long prison sentences in an environment designed to be punitive rather than rehabilitative; some crimes might even be punishable by death.

Should we have Harsh Retributivism? Can this be justified? According to *The Epistemic Argument*, versions of which have been defended by Gregg Caruso, Ben Vilhauer, Elizabeth Shaw and me, the answer is “no”; we can’t be *sufficiently certain* that we do the right thing when harshly punishing offenders.

The epistemic argument

There are plenty of arguments that aim to show that Harsh Retributivism is *wrong*. One might argue that harsh retributivism breeds *more* crime, and we should consider keeping the crime rate low the *most* important job of the criminal justice system, for the sake of future potential victims. One might argue that free will is a myth, and therefore, no one can deserve to be punished. Epistemic arguments against Harsh Retributivism differ from others in that they don’t try to establish that Harsh Retributivism is *definitely wrong*; merely that we can’t be sufficiently certain that it’s *right*.

To intentionally harm someone else, to intentionally cause other people suffering, is serious business. That’s not to say that doing so is *always* wrong, but the burden of justification is strong.

This is something we should all agree on. Sometimes people quibble about what *counts* as harming – for instance, whether taxation harms the persons taxed or not. A colleague of mine claims that it’s a minor harm that can nevertheless be justified by the benefits, whereas I don’t think it counts as a harm at all (roughly, my arguments are that I wouldn’t be better off in either a taxation-free, libertarian nation, nor in a situation where I kept cheating on the taxes and getting away with it; the latter would have drawbacks that outweighed the economic benefits).

Nevertheless, regardless of our stance on something like taxation, we should all agree that being in a harsh, punitive prison environment for a long time harms a person. Thus, we shouldn’t do this to people unless we’re highly certain that it is the right thing to do. This is acknowledged when criminal courts demand that no one should be convicted unless there’s certainty beyond a reasonable doubt that they committed the crime, but it should also be acknowledged when we discuss whether the system as a whole can be justified.

For Harsh Retributivism to be morally right, at least these three statements must be true:

1. People in general have free will and moral responsibility for what they do, making it possible in the first place for someone to deserve something.
2. Retributivism is the correct criminal justice theory; the criminal justice system's proper job is to deal out just deserts.
3. We know how to match up crimes and punishments, and the fitting punishment for many crimes is a *harsh* one.

All three statements can be argued for. Many philosophers argue, for instance, that “free will” is not a mysterious, near-supernatural thing, but should be conceived of as a set of fairly simple, down-to-earth psychological capacities that most people have. Many of those philosophers further argue that this is sufficient for us to deserve praise and blame, rewards and punishments for what we do. Some philosophers argue that only a retributivist criminal justice system takes the dignity of victims seriously. Furthermore, some say, we should trust the judgment of virtuous people; a virtuous person feels that both themselves and others should be made to suffer in return, if they had made others suffer through their crimes.

Disagreement should make us less certain

Now, the epistemic argument against harsh retributivism doesn't require us to prove that at least one of the three statements above is *false*. It doesn't require us to point out definite holes in crucial arguments for the three statements. Instead, the epistemic argument shows that we can't be *certain enough*. Even if we initially feel certain about our views because we feel certain about our arguments, the fact that so many philosophers and scholars disagree with us should prompt us to dial down our certainty a fair bit.

There are philosophers who argue that ordinary psychological capacities *cannot* make us deserve punishment – in order to deserve anything, we would need an impossible kind of free will. There are philosophers who argue, often independently of metaphysical free will questions, that the proper justification for the criminal justice system is something other than retributivism. And even retributivists, as we have seen, might argue that people do not deserve *harsh* punishments. We have our arguments, but they have theirs, and we have no independent basis for claiming that our philosophical opponents are any less intelligent, rational, well-read, or conscientious in their argumentation than we are.

It's generally a sound principle to dial down one's initial certainty when you find out that equally smart people are of a different opinion. To take a simple, everyday example: Suppose

I have dinner with friends at a nice restaurant, and we decide to split the bill. I'm generally good at math and calculating sums in my head, so I look at the bill and announce that we should each pay 45 dollars. Initially, I feel certain, but when a friend says that it's 47 dollars, I should immediately dial down my certainty – at least if I have no reason to think that they're any worse at math than I am.

Of course, this kind of uncertainty is easily resolved – one of us need only pull up their phone and use a calculator app. But if, for some reason, we had calculated different sums in our heads and had neither devices nor pen and paper with which to check who's right, and if repeated head calculations still give different results, the reasonable thing to do would be to admit that neither of us know for sure. Perhaps my friend made a mistake, but then again, perhaps I did.

In all the philosophical debates above, we have no way to check, once and for all, who's right and who's wrong. All the arguments are out there in the open. Sceptics of moral responsibility and desert have read the arguments of the desert defenders, and the other way around. All the camps in the criminal justice debate are familiar with each other's arguments and positions. Still, they differ in their judgments of *how strong* various arguments are. For my own part, I think (perhaps pessimistically) that there are many philosophical issues on which we will never reach agreement, because the arguments ultimately bottom out in intuitions – that something just *feels* right or something just *feels* wrong – and those intuitions differ from person to person. In the free will and moral responsibility debate, for instance, there's broad agreement that people *do* have certain psychological capacities – we can weigh our reasons, consider what to do and act on our decisions – and also that we *don't* have any mysterious, supernatural abilities to initiate new actions independently of our brains, or independently of the laws of nature. But what follows from this? Can we still deserve things, or can't we? This last question moves us outside the realm of empirical science – there is, after all, no “deservometer” with which we could measure the presence or absence of desert – and into philosophy proper. And however clear and rational people's arguments, here and there appeals to what just *seems* to be the case with regards to desert will pop up.

The same problem comes up when discussing what the criminal justice system's primary purpose ought to be, and, if we take retributivism as a premise in the discussion, which punishment fits which crime. At the end of the day, people's intuitions differ. It's plain irrational to take for granted that my intuitions are the best, and everyone else's flawed.

At this point, someone might ask whether this uncertainty isn't an equally big problem for alternative views on criminal justice. Not quite. Compare the following two situations:

1. We have a humane, restoration-and rehabilitation-focused criminal justice system. Harsh retributivists are unhappy about it, because they think that criminals don't get *as much* punishment as they deserve.
2. We have a harsh and punitive criminal justice system. Deterrence theorists, rehabilitation theorists, restorative justice advocates, soft expressivists and soft retributivists are all unhappy about it, both because they think that other important values get ignored, and because they think that criminals get *undeserved* harsh punishments.

1 and 2 are not symmetrical, even though some people are unhappy about the situation in both. Everyone should agree that it's worse to deal out undeserved punishments than to give someone less punishment than they deserve (or none). If too little punishment *were* as bad as too much, the reasonable thing to do would be to convict people accused of crimes as soon as it seemed a bit more likely than not that they did it, but a court system operating by that principle would be a nightmare.

We should err on the side of caution. That means that we should both keep the presumption of innocence and demand proof beyond a reasonable doubt in individual court cases, *and* reform harsh retributivist criminal justice systems in a preventive, restorative, rehabilitative and humane direction.