THE ONGOING REVOLUTION IN PUNISHMENT THEORY: DOING JUSTICE AS CONTROLLING CRIME

Paul H. Robinson*

You probably remember from your first-year criminal law class the age-old tension between the retributivists who want to punish offenders because they deserve it – they see deserved punishment as a value in itself – and the utilitarians (or instrumentalists), who believe that punishment must have some more practical justification such as avoiding future crime, perhaps through deterrence, incapacitation of the dangerous, or rehabilitation. The dispute between these two groups is classically thought to be irresolvable. The two are simply using different currencies, and think different things to be important.

One of the most exciting developments in current punishment theory suggests that these two positions may not be entirely irresolvable, at least in a sense. You all know the American Law Institute and the Model Penal Code that it promulgated in 1962. Almost three-quarters of the states have since codified their criminal law in ways modeled after that code. Last year, for the first time in 46 years, the Model Penal Code was amended to change the section setting out the purposes of the code, its provisions, and how those provisions are to be interpreted. And much of what I will talk about today is the story – in intellectual terms – of how that change in perspective came about.

I'll come back to the Model Penal Code amendment at the end of the talk. Let me start with some background leading up to those recent events.

The most fundamental question for criminal law may be this: how can we justify having a system that imposes punishment? The moral philosophers have killed many forests answering this question but, to be honest, I've always found it a bit boring. It seems to be a purely academic debate. Every alternative purpose on the list of possibilities argues for the same conclusion: yes, we should have a criminal justice system that imposes punishment. Such a system would be useful because it provides the opportunity to impose deserved punishment, to deter future offenses, to take control of dangerous persons, to rehabilitate offenders that need it, and so forth. It is no surprise that every known organized society has had some kind of system of criminal justice that imposes punishment on wrongdoers.

The more difficult, but much more interesting question, is this: If we are to have a criminal justice system, who is to be punished by that system, and how much? This is not an academic question; it is in fact the core of what criminal law and theory is about. It is a question of enormous practical significance. What kind of instructions should a legislature give to a criminal code revision commission? What kind of instructions should be given to a developer of sentencing guidelines such as the United States Sentencing Commission, created by the Sentencing Reform Act of 1984? What sort of interpretive rules should a criminal code give to

---

* Colin S. Diver Professor of Law, University of Pennsylvania. This is the text of the First Annual Edward J. Shoen Lecture, delivered at the Sandra Day O'Connor College of Law, Arizona State University, on February 26, 2009. The lecture is drawn from PAUL H. ROBINSON, DISTRIBUTIVE PRINCIPLES OF CRIMINAL LAW: WHO SHOULD BE PUNISHED HOW MUCH? (Oxford 2008).

judges in applying that code? What sort of rules should sentencing judges use in exercising their discretion?

Unless we are to allow every individual commissioner and every judge to just make things up as they go along (which would be obviously and seriously problematic), we must provide some sort of guidance. If a commissioner or judge must decide whether to adopt any given formulation of a rule, or to adopt any particular interpretation or to impose a particular sentence that advances one purpose at the expense of another, what kind of rational guidance can they be given in making that selection?[1]

These are particularly challenging questions in part because, of the alternative distributive principles that one might adopt to decide these issues, different principles will distribute liability and punishment in different ways than will any other.[2] Thus, the central question in criminal law theory – the question I want to talk about today – is: Which of the alternative principles (or which combination of principles) for distributing criminal liability and punishment should we rely upon?

First, for terminological purposes, let me list what I mean by each of the alternative principles that I will mention.

[Insert 1]

Each of these alternative distributive principles looks to a different criterion and therefore would distribute different amounts of liability and punishment to different people in different situations. Consider, for example, the case of the elderly man who was once a Nazi torturer. If one is focusing purely on *deserved* punishment, this would be an easy case for imposing liability because what the person has done is so horrible. Similarly, if your focus is general deterrence – using this case as an opportunity to dissuade others by the threat of liability and punishment for engaging in such conduct in the future – then again, there is good reason to punish. On the other hand, if one's focus is incapacitation or rehabilitation, punishment in such a case may be a complete waste of resources. The offender may no longer be a danger. Or, if the focus is special deterrence – imposing punishment in order to deter future offenses by the offender at hand – punishment again may be a waste of resources; this offender is not likely to ever again have the opportunity to engage in such horrors.

Consider too the question of whether the criminal law ought to recognize an excuse defense for people who are seriously mentally ill and have, because of that illness, committed an offense without realizing its wrongfulness. If the focus is desert, there is no blameworthiness and therefore no reason to punish. Similarly, if the focus is special deterrence, such people may well be sufficiently dysfunctional that they simply do not understand, or cannot respond in a rational way to, the deterrent threat. On the other hand, there may be very good reasons to deny an excuse defense and impose liability on the seriously mentally ill offender if one's distributive principle is incapacitation or rehabilitation. The person has already demonstrated that their mental illness makes them dangerous. Finally, if the distributive principle were general deterrence, liability in such cases could be an effective way of signaling all non-insane potential

---

[1] For a more detailed discussion, *see id.* at ch. 1.

[2] For a more detailed discussion, *see id.* at ch. 2.

offenders that the criminal law is quite serious about punishing this kind of conduct, and that they ought not hold out some hope that they can commit the offense but find some insanity loophole. Indeed, denying an insanity defense can present a useful opportunity to reinforce the seriousness of the deterrent threat.

Notice, most interestingly, that not only do different distributive principles answer the insanity defense question differently, but that the alignment of principles on each side of the question is different for the issue of the insanity defense than it is for the previous issue of what to do with the elderly Nazi torturer. Other questions will similarly generate altogether different alignments.

Consider, for example, the question before a criminal code commission that must decide whether to have attempts punished the same as the completed offense or to have attempts punished less. A distributive principle of incapacitation or rehabilitation would suggest that they be punished the same, at least in the common case in which the attempter and the successful offender are equally dangerous. Yet general and special deterrence and at least some forms of desert would suggest that the attempter be punished less.

In other words, the distributive principle we adopt for the criminal justice system really does make a difference. We must make some choices. My plan for the next 40 minutes is to work through each one of the alternative distributive principles and to evaluate the strengths and weaknesses of each in the hope that, at the end of the process, we can step back, evaluate the alternatives, and make some judgment about which principle might seem most attractive (or at least decide which is the least unattractive).

To give a hint at where we are headed, it turns out that none of the alternative distributive principles are perfect. All have flaws. It will be a complicated decision at the end to pick one, or to try to construct some kind of hybrid principle that combines two or more. That task will require us to have a somewhat nuanced appreciation for the relative strengths and weaknesses of the different distributive principles, so let us consider each in turn.

DETERRENCE

The most obvious strength of general deterrence is its enormous potential efficiency. By punishing a few people, there is the potential for influencing the conduct of thousands or hundreds of thousands of other potential offenders.[3]

On the other hand, however, there can be no deterrent effect unless certain basic prerequisites are satisfied. First, the target must know the rule by which we seek to influence the person's conduct. Second, the target must be able and willing to calculate the costs and benefits for their own self-interest – that is, they must be able and willing to take the threat contained in the deterrence-based rule and use it to decide how they will act. Third, having done those calculations, the target must come to the conclusion that the costs of the contemplated offense, in terms of punishment, outweigh the benefits of the offense. Yet, as I'll discuss, these prerequisites commonly do not exist. Situations in which they all exist (a necessary prerequisite for any deterrent effect) are the exception rather than the rule.

Let's look at each of the three prerequisites in turn, to see the real world complications that arise in trying to satisfy each. As to the first prerequisite – that the target must know the rule

[3] For a more detailed discussion, *see id.* at chs. 3 & 4.

by which we seek to manipulate his conduct – the studies suggest that offenders commonly don't know the rule. How many people here know what the Arizona rule is on felony murder? On "three strikes"? Not many – and most the people here are legally trained. What are the chances that a young man standing outside of a convenience store contemplating a robbery will know the rule? Yet it is this person whose conduct the rule seeks to influence.

One thing that helps us in this instance is that the more involved a person is in the business of criminality, the greater their incentive to actually learn the criminal law's rules. But what we know from studies is that even criminals commonly have no idea what the rule is and, even if they think they know, they often have it wrong.

Even if the target knows the rule, there can be no deterrent effect unless the person is able to take that information and use it to rationally guide their conduct. The problem here is that the people most likely to be offenders are the people who are most likely to be bad calculators, or be indifferent to future consequences. They are more likely to be under the influence of alcohol, drugs, anger, fear, group arousal, group identity shift, over-impulsiveness, mental illness, or a variety of other factors making them dysfunctional in their calculations. This is not true of every offender, of course. White collar offenders, for example, are much more likely to be the rational calculators that deterrence requires.

Even if the target knows the rule and is a rational calculator, however, there can be no deterrent effect unless the target perceives that the costs of the contemplated offense outweigh its benefits. As Bentham has famously suggested, this perception is the product of three things: the probability of punishment, its intensity (its amount), and its timing (the delay of the punishment following the offense). Let's look at each of these components of the deterrence formula in turn.

The effect of probability of punishment on deterrent value gives us some reason to be concerned. While we can't make too much of animal studies – I'll cite a few different ones during my talk today – they can be useful in at least giving a hint at where potential problems may be, because many species share the physiological reactions of humans. So, for example, studies of rats suggest that deterrent effect drops off fairly quickly as the punishment rate decreases. Where most offenses have a punishment rate of 100 or 50 to 1, it should not be complete surprise that this threat level proves inadequate to deter.

We are helped here by the fact that people tend to exaggerate the likelihood of rare events. That's why people buy lottery tickets. However, the more a person is inclined toward criminality, the more they may have a special incentive to sort out what the real rate of punishment is, and thus may discover that is quite low.

That people tend to exaggerate their own abilities also hurts deterrence. That is, an offender may hear that another burglar has been caught and punished, but may discount the significance of this to him on grounds that he is a much better burglar than the other person. The unfortunate bottom line is found in studies like that by Anderson, who suggests that many if not most offenders think the threat of punishment sufficiently low as to be of limited importance to them.

Regarding the amount-of-punishment part of the equation, it seems clear that we can, of course, impose punitive bite. What is less clear is whether we can modulate the amount of punishment, as a deterrence-based system would require. There are a host of facts that make it difficult to predict and control the amount of punishment that will be felt and remembered.

One well-known complication is found in the human tendency toward discounting future benefits and costs. Thus, most people would prefer to take $100 in hand right now rather than an

absolute guarantee of $150 a year from now, even though they can't earn 50 percent annual interest.

Another complication comes from what one might call "learning the pain." Having trained rats to get food by pressing a bar, one can deter them from continuing to press the bar by applying an 80 volt shock. However, the application of a 60 volt shock is not likely to deter. The animal will simply take the pain to gain the food. If one gives a 60 volt shock (which the animal will accept), then a 70 volt shock (which the animal will also accept), *then* the 80 volt shock, the animal will not be deterred, because the two lower-intensity shocks have taken away the value of the higher punishment. The animal learns that he can indeed take the 60 volts, and the 70 volts; thus the 80 volts no longer holds the intimidation value it had when applied as the starting point. Indeed, when increasing incrementally, the animal takes up to 300 volts before being deterred! The point here is that the dynamic of deterrence can be a complex one.

Another source of complexity is shown by subjective well-being studies. Whether a person wins the lottery or becomes a paraplegic in a car accident, it is not uncommon for the person to adjust their baseline perspective over time. The lottery winner is ecstatic for while, but may return to their old cranky, unhappy self as their expectations adjust upwards. The same can happen in reverse for paraplegics. This is not good news for a deterrence regime, for it means that the bite of imprisonment is likely to dissipate over time, making each successive unit of imprisonment less painful even though the cost of each unit remains the same. This produces an increasing inefficiency as the permitted prison term gets longer. The bar graphs below illustrate the point. While policy makers tend to think in terms of the "naive calculation" in Bar 1, the reality is closer to the "adaptation calculation" in Bar 2.

[Insert 2]

In fact, human nature makes things even worse for deterrence. "Duration neglect" studies suggest that it is both the maximum intensity and the endpoint intensity that determine the remembered punitive bite, but that the duration of punishment is of relatively negligible effect. This is particularly significant when combined with the point just noted – that the punitive bite of imprisonment naturally degrades over time. The longer the term of imprisonment, the lower the perceived endpoint intensity. Thus, at least theoretically, longer terms of imprisonment such as in Bar 2 below may be remembered as having less total punitive bite than the shorter term of imprisonment in Bar 3, because the important endpoint intensity dissipates over time in Bar 1.

[Insert 3]

Regarding the last component of the deterrence formula – the delay in punishment – consider another animal study. Hungry dogs are given separate bowls containing good tasting food and bad tasting food. At a session after they have sorted out which is the good tasting food, they are allowed back into test room and immediately go for the good tasting food, but then are bonked on the nose with a newspaper by the experimenter (bonking being an experience that dogs do not like). Some dogs are bonked immediately, some after a 5-second delay, and some after a 15-second delay. When the dog is later allowed back into the room, we can see how the different delay periods affect the length of time it takes them to return to the good tasting food. The dogs bonked after 15-second delay take three minutes to try eating the good tasting food

again. In contrast, those dogs bonked more quickly, after only a 5-second delay, take eight days before they risk eating the good tasting food again! Those bonked immediately take two weeks before they try again. Clearly, the timing of punishment can be highly influential on deterrent effect.

Given that in the criminal justice system punishment is almost always long delayed – on average seven months after a guilty plea and 13 months after trial and conviction – one may worry that the delay can seriously undermine the deterrent effect. Again, the point here is simply that deterrence is a complex business with complex dynamics, and there are many ways in which deterrent effect can be lost.

Compare these difficulties in establishing the costs of crime – low probability and long delay – with the perceived benefits of a crime, which are likely to be perceived as having a high probability (because of the low capture rate) and an immediate tangible benefit.

This skepticism about the satisfaction of the three prerequisites for deterrence is compounded by additional independent concerns about complexity in the application of a deterrence program. The problem of substitution effects, for example, is well known. Even if a deterrence program is successful, it might simply drive offenders to commit more serious offenses, as Katyal has noted. Adding to the problem are dynamic effects. By being successful in creating a deterrent effect, one is likely to necessarily alter the parameters of the deterrence equation for the future, as Bar-Gill and Harel have suggested.

Ultimately, the difficulties of deterrence are shown in the aggregated-effect studies, which use real-world experiments – such as gauging crime rates before and after a rule or policy change – to determine whether the rule or policy has had the desired deterrent effect. The studies are useful because even if it seems that the prerequisites for effective deterrence are not present, it might be that there nonetheless may be a deterrent effect that somehow operates in ways that social scientists do not yet understand.

But the studies suggest just the opposite. That is, they seem to bear out the conclusion that effective deterrence is possible only if the prerequisites are satisfied, but that the prerequisites commonly are *not* satisfied. The studies show that even when reformers actively seek to increase deterrence, there is most commonly no effect, and where there is any effect it is at best near trivial and often unpredictable, sometimes increasing a particular sort of crime rather than reducing it.

This is not to say that deterrence can never be increased through the manipulation of legal rules. Under special circumstances – where all three prerequisites are satisfied – one can produce a modest deterrent effect, albeit one that can quickly fade. The ultimate conclusion, then, is not to reject deterrence as a distributive principle but rather to understand that it has specific requirements and has limited situations in which it can be effective. Its strengths and weaknesses might be summarized this way. (I'll provide such a summary for each alterative distributive principle to make it easier to compare them in a single table at the end.)

[Insert 4]

REHABILITATION

There are reasons to be skeptical about rehabilitation as a distributive principle for criminal liability and punishment.[4] First, such programs typically work only for limited kinds of offenses and offenders. When they do work, the effects tend to be quite modest. Perhaps even more problematic, if rehabilitation were the sole distributive principle for liability and punishment, the system would release all offenders who could not be rehabilitated – which would probably be the majority of those incarcerated today. Needless to say, this would likely be seen as unworkable by anyone, whether their interest is in crime control or in doing justice. It probably explains why no known system has ever used rehabilitation as its sole distributive principle. We will consider later whether rehabilitation might be used as part of a hybrid distributive principle in conjunction with some other principal. For example, one might construct a hybrid distributive principle that relied upon rehabilitation but, if rehabilitation is not possible, reverts to incapacitation.

On the other hand, rehabilitation might be a very good correctional policy. That is, one might want prison wardens and probation and parole officials to use rehabilitation programs as much as possible for those offenses and offenders were rehabilitation is, in fact, possible. Used under the right circumstances, it could well be a good investment. What is problematic is using rehabilitation as the basis for deciding who the criminal justice system should restrain and for how long.

One last point is worth mentioning. There may well be special value to rehabilitation apart from its crime control potential. Most would agree that there is intrinsic value in helping people to change themselves to live more useful and satisfying lives to their full potential.

INCAPACITATION OF THE DANGEROUS

Unlike deterrence and rehabilitation, there is no doubt that incapacitation can effectively control future crime. It does work. And I think it equally clear that there is general agreement that society has a right to protect itself by restraining dangerous persons, as it does now with the civil commitment of those who are mentally ill, who have contagious diseases, or who have a drug dependency.[5]

The weakness of incapacitation comes from its use as a distributive principle for criminal liability and punishment within the criminal justice system. The problem is that the system, for good reasons that I'll talk about, continues to advertise itself as being in the "criminal justice" business, as if it was designed solely to punish people for past offenses. Thus, when the system is being used for preventive detention purposes – incapacitating persons in order to prevent future crimes that we think they may commit – it must "cloak" its preventive nature; it must continue to pretend that it is only bringing about justice for a past offense. And it is that preventive-detention use of the criminal justice system, with its attendant "cloaking," that produces serious problems. To telegraph my conclusions: Such a practice undermines the system's predictive accuracy, which in turn reduces its preventive effectiveness and efficiency.

---

[4] For a more detailed discussion, *see id.* at ch. 5.

[5] For a more detailed discussion, *see id.* at ch. 6.

At the same time, it increases the criminal justice system's unfairness to detainees, and undermines the criminal justice system's moral credibility with the community, which in turn reduces its normative crime control effectiveness.

One may wonder: why bother doing preventive detention within the criminal justice system? Why not do it as part of a separate, open, civil, preventive detention system? There are good reasons why not; system designers do in fact have reasons for being attracted to preventive detention cloaked as criminal justice, rather than to open preventive detention.

The older people in the audience may remember the political upheavals the 1970s that surrounded what were rather modest proposals for open preventive detention, such as detention after arrest pending trial. These battles produced enormous political pressure. By cloaking preventive detention as criminal justice, system designers can get where they want to go without this "heat" on their backs.

The cloaking also allows the system to avoid the limitations that logically follow preventive detention, such as the legal limitations that we put on civil commitment. More on this in a moment.

Finally, cloaking preventive detention as criminal justice also has the benefit of reducing the visibility of the criminal justice system's shift away from deserved punishment. This is important because greater openness about this shift would tend to undermine the criminal justice system's "moral credibility," and thereby undermine its normative crime control effectiveness. I'll discuss this in more detail when we take up desert as a distributive principle.

While one can understand why such cloaking is attractive to system designers, the fact is that such cloaking is also seriously problematic for the effectiveness of preventive detention. Let me give two examples.

Where the goal is incapacitation of dangerous offenders through the criminal justice system, it is common to use a person's criminal history, rather than a direct quantification of their current harmfulness, as the measure of their dangerousness. This is because the former appears to be related in some way to the offender's blameworthiness, and thus looks more like justice. That is, it makes the criminal justice system look more like it is doing criminal justice and less like it is doing preventive detention. Unfortunately, criminal history is a significantly less accurate predictor of future criminality than would be a direct clinical assessment of a person's dangerousness.

To give another example, when preventive detention is being done within the criminal justice system it commonly uses determinate sentences, rather than release decisions contemporaneous (or nearly so) with release. That is, judges impose a sentence, including the release date, at the time of sentencing rather than delaying the decision until the time when release is contemplated. For cloaking purposes, this practice makes good sense: all of the factors affecting deserved punishment are available at the time of sentencing. If preventive detention is to be cloaked as doing justice, there is little justification for delaying the release decision.

The difficulty with this cloaking practice is that making release decisions at the time of sentencing significantly reduces their accuracy for preventive detention purposes. It is difficult enough to predict future criminality at the time of release. It is much more difficult to accurately predict future criminality months or years or decades earlier, when the offender is initially being sentenced.

These kind of cloaking practices are bad for effective and efficient preventive detention because by reducing predictive accuracy, they provide less protection at greater cost. Relying

upon criminal history as the stand-in for dangerousness, for example, means that dangerous persons can be released because of insufficient criminal history. Consider the case of stalking as a first conviction. If social scientists can reliably predict that the stalking will mature into greater violence, good prevention would favor detention – but the lack of criminal history would translate into immediate release.

At the same time, relying upon criminal history as the stand-in for dangerousness means that non-dangerous persons will be detained simply because they have a criminal history or because they were dangerous back when they were sentenced. The problem is apparent in the operation of the "three strikes" rule. The common trajectory for violent offenders is increased violence in the late teens and 20s – the testosterone years – that trails off as the offender gets older. Thus, the focus on criminal history under "three strikes" means that the young offender runs free during his most violent time, as he is building his criminal history, then is imprisoned after he has finally reached his third strike – often as his violence is naturally tapering off. This, unfortunately, tends toward producing a collection of geriatric nonviolent detainees.

Not only is preventive detention cloaked as criminal justice bad for effective preventive detention, it is also unfair to detainees. By doing preventive detention under the guise of criminal justice, it avoids the limitations that logically should constrain it. Specifically, where the justification for detention is the prediction of future criminality, the system logically ought to have a principle of minimum restraint. That is, the preventive detention rationale justifies no more restraint than is necessary to protect society, which may mean the use of methods other than imprisonment. Preventive detention also logically means nonpunitive conditions, as is now the case with civil commitment of the mentally ill. It also means a right to treatment, if that treatment can reduce the need for further detention. It also means a right to periodic review. That is, the government ought to have to regularly show continuing dangerousness if it seeks to continue preventative detention. Finally, a preventive detention system logically would set some minimum decision standards – requiring, for example, some defined minimum likelihood of a violation, some minimum defined level of harm seriousness threatened, and some minimum defined level of predictive reliability.

By using incapacitation as a distributive principle for criminal liability and punishment, all of these logical limitations called for by preventive detention system are avoided. If we are to have preventive detention – and there are good arguments for why society ought to be able to protect itself with such a system – it would be better for both society and for detainees to have that preventive detention done openly through a system of civil commitment rather than cloaked as a system of criminal justice.

I understand the resistance of many: that permitting such an expansion of civil commitment system would be dangerous. This has been discussed as what one might call the "Gulag Archipelago" Problem. Liberal democracies, and American values in particular, naturally want to put significant limits on the power of government to intrude in the lives of its citizens. Punishing criminals is one thing, but detaining people based only on a prediction of future criminality is another.

However, the important point here is that our current practice of doing preventive detention cloaked as criminal justice creates the potential for greater abuse, not less. It avoids the scrutiny that an open preventive detention system would attract. It avoids the logical constraints to which preventive detention should be subject.

If you have any doubt about this, imagine the *Rummel* case coming before the Supreme Court as a preventive detention case rather than as a criminal justice case. Rummel was convicted of a third minor fraud and under a three strikes statutes was sentenced to life imprisonment. The Supreme Court held the sentence constitutional. There are many good reasons not to constitutionally constrain the criminal justice sentencing process too tightly; criminal justice does sometimes need life imprisonment as a punishment. But if Rummel had come before the Court as a preventive detention case, which in reality its three-strikes predicate made it, one can imagine that it would not have passed the Court's laugh test. Automatic life detention to avoid another minor fraud offense? I don't think so.

The strengths and weaknesses of rehabilitation and incapacitation might be summarized this way:

[Insert 5]

EMPIRICAL AND DEONTOLOGICAL DESERT

I noted earlier that desert as a distributive principle would give criminal liability and punishment according to an offender's blameworthiness, which would take account of the extent of the harm or evil of his conduct, his culpable state of mind at the time of the offense, an assessment of his personal capacities that might shape what we could reasonably have expected of him, and a variety of other factors.

I have distinguished here the traditional deontological desert from what has been called "empirical desert." The former is an assessment of moral blameworthiness logically derived from principles of right and good, typically by moral philosophers. The latter is derived from social science studies of a community's shared intuitions of justice. Empirical desert is not "true justice" in a transcendent sense, but only a representation of the principles by which the community actually makes judgments about justice.[6]

It is obvious why one might care about doing justice in a deontological sense. Why might one care about empirical desert? Why might empirical desert be an attractive distributive principle for criminal liability and punishment? The short answer, we will say, is that it might make sense for instrumentalist crime control reasons. We will come back to that revolutionary idea – that doing justice, at least in an empirical desert sense, might be an effective crime control strategy – because it does put a new spin on the traditional view that the retributivist interest in doing justice and the instrumentalist interest in controlling crime are inevitably in conflict.

To work up to this conclusion, let me give some background from the social science studies of the past decade or more.[7] Laypersons see punishment as something that is properly imposed according to desert; that is, blameworthiness. When they are asked to assign punishment, they don't look to the factors that determine dangerousness or deterrence, but rather to the offender's moral blameworthiness.

As discussed earlier, the traditional instrumentalist crime-control principles of deterrence, incapacitation, and rehabilitation each conflict with a desert distribution of punishment. If any of

---

[6] For a more detailed discussion, *see* ROBINSON, *supra* note 1, at ch. 7.

[7] For a more detailed discussion, *see id.* at ch. 8.

these principles were used for distributing criminal liability and punishment, the system would regularly do injustice, and would often fail to do justice. Conversely, if the system adopted a desert distribution, it would not be optimizing deterrence, incapacitation of the dangerous, or rehabilitation.

However, a desert distribution of criminal liability and punishment *would provide some significant opportunity for* deterrence, incapacitation of the dangerous, and rehabilitation, albeit not the maximum that is possible. That is, deserved punishment can have a deterrent effect, can incapacitate, and can provide the opportunity for rehabilitation. The important point here is that to increase any of these instrumentalist effects, the distribution of criminal liability and punishment must deviate from desert – that is, it must do injustice or must fail to do justice

Do crime-control instrumentalists have any reason to care about whether the criminal justice system regularly does injustice or fails to do justice? As instrumentalists, deviating from true justice (deontological desert) may be just an unfortunate necessity of fighting crime, one might argue. However, social science hints that there may be practical real-world crime-control complications that arise from regularly deviating from the community's perceptions of justice (that is, from empirical desert).

Here's how. We are becoming increasingly aware of the enormous power of social influence and internalized norms. The behavioral decisions that people are constantly making in their daily lives are driven primarily by a concern for what others, especially family and friends, will think of them and for what they think of themselves. The criminal law can harness these normative forces if it earns a reputation as a moral authority – that is, if people come to see it as a system that reliably punishes in ways consistent with people's intuitions of justice.

A criminal justice system that has earned moral credibility within the community is in a position to harness the power of stigmatization, for example – a highly efficient mechanism for influencing conduct. It lacks the high costs of imprisonment, yet can significantly influence people's conduct. In contrast, if the criminal law fritters away its moral authority by imposing criminal liability and punishment that deviates from empirical desert, it increasingly undermines its ability to stigmatize conduct through criminalization or punishment.

A criminal law that has earned a reputation as a moral authority also has a greater ability to avoid vigilantism, which is classically sparked when the community sees regular failures of justice that it finds intolerable. Similarly, a criminal justice system that regularly does injustice and or fails to do justice is one that risks prompting resistance and subversion, and loses its ability to gain the acquiescence and cooperation that a criminal justice system relies upon, by witnesses, jurors, offenders, and most participants in the criminal justice and correctional process.

The criminal justice system that has earned moral authority also has a greater chance of gaining compliance in borderline cases where the actual condemnability of the conduct may be unclear. When insider trading first became a crime, for example, it may not have been immediately obvious to everyone that this conduct was qualitatively different from other forms of aggressive entrepreneurship that are tolerated and even encouraged. If the criminal justice system has earned a reputation as a reliable guide to what is and is not condemnable conduct, it is more likely to gain the deference of the community when it announces that insider trading has crossed a line and is indeed condemnable.

Perhaps the most powerful effect of gaining moral credibility is the influence that such credibility gives to the system in the larger public conversation by which societal norms are

shaped. If we want to change people's thinking about the condemnability of domestic violence, or drunk driving, or downloading music from the Internet without a license, criminalization of that conduct or increasing the penalty to signal greater seriousness of the conduct can help reinforce the norm against it. In contrast, a criminal justice system that has squandered its moral authority by regularly deviating from desert is one that is more likely to be ignored during the public conversation, because its view may be discounted as just one more example of how the system gets it wrong. (Understand that any criminal law, even one with moral credibility, may not be able to establish a strong societal norm by itself. That is the lesson of Prohibition. A strong, and eventually internalized, norm requires concurring views from a variety of sources of moral authority, including social institutions as well as circles of friends and acquaintances.)

One may conclude, then, that the crime-control power of the criminal law depends in some significant part upon how well it tracks the community's shared intuitions of justice. Thus, let me say a few words of background about lay intuitions of justice. First, we have learned that people's intuitions of justice are quite nuanced and sophisticated. Small changes in facts can and do produce large and predictable changes in liability judgments. And sophistication does not depend upon people's education or intelligence; it seems to be the standard form.

Even more surprisingly, there appears to be an enormous amount of agreement about intuitions of justice across all demographics, at least with regard to the core of wrongdoing – physical aggression, taking property without consent, and deceit in exchanges. The agreement here is on the relative blameworthiness of different kinds of offenses and offenders, not on the absolute amount of punishment to be imposed. However, once a society commits itself to a punishment continuum endpoint, as every society must do (whether it is the death penalty or life imprisonment or 20 years), the large number of cases of distinguishable blameworthiness must be fit on this limited punishment continuum. Thus, each case will end up requiring a specific amount of punishment, not because of any magical connection between that amount of punishment and that offense but rather because that specific amount of punishment is required to put that case in its proper ordinal rank among all other cases. (If one changes the punishment continuum endpoint – different societies do have quite different endpoints – then the specific punishment required to put each case its proper ordinal rank would also change.)

One may well ask how well current American criminal law matches the community's intuitions of justice. The short answer is: not well. Modern crime-control programs, such as three strikes, high drug offense penalties, adult prosecution of juveniles, narrowing the insanity defense, strict liability offenses, and the felony-murder rule, all distribute criminal liability and punishment in ways that seriously conflict with lay persons' intuitions of justice.

To summarize, then, the strengths and weaknesses of the empirical and deontological desert might be presented this way:

[Insert 6]

SELECTING A DISTRIBUTIVE PRINCIPLE

We have now considered each of the alternative distributive principles, and it is time to answer the question with which we started: What should be the principle by which we distribute criminal liability and punishment? Here are our summaries collected in a single table:

[Insert 7]

All of the alternative distributive principles are flawed in one way or another. Are we compelled to adopt the least flawed of the group? Or, could we combine two or more distributive principles in one way or another to create a hybrid?[8]

The original "purposes" section of the Model Penal Code provided a hybrid of sorts: a laundry list of all of the alternatives, without any articulation as to how they were related to one another.[9] The problem was that when they conflicted with one another – which we know happens regularly – the Code's provision provided no guidance as to how that conflict was to be resolved. Worse, by seeming to provide a means of principled decision making, it created a façade that hid the real potential for abuse: A decisionmaker could simply reverse-engineer the process and announce a "principled" decision, by simply deciding what result he wanted, for whatever unarticulated reason, then looking to the laundry list of alternative distributive principles to determine which one would give that result; he could then announce a decision "based upon" that selected distributive principle.

A principled system must look to a distributive principle that defines the interrelation among alternatives. In its recent amendment, the American Law Institute adopted a revised "purposes" provision that did just this. Here's the text of the new provision:

> (2) The general purposes of the provisions governing the sentencing and corrections, to be discharged by the many official actors within the sentencing and corrections system, are:
>
> (a) in decisions affecting the sentencing and correction of individual offenders:
>
> (i) to render punishment within a range of severity ***proportionate to the gravity of offenses, the harms done to crime victims, and the blameworthiness of offenders***;
>
> (ii) when possible with realistic prospect of success, to serve goals of ***offender rehabilitation, general deterrence, incapacitation of dangerous offenders, and restoration of crime victims and communities***, provided that these goals are pursued within the boundaries of sentence severity permitted in subsection (a)(i); and
>
> (iii) to render sentences no more severe than necessary to achieve the applicable purposes from subsections (a)(i) and (ii); . . . .[10]

As you see, the A.L.I. has adopted desert as the Model Code's dominant distributive principle. One may speculate that it came to this conclusion for the kinds of reasons reflected in the analysis of the strengths and weaknesses of alternative principles that we have just worked through.

As hinted at above, this position seems to resolve the traditional retributivist – instrumentalist "irresolvable tension" – but only in a sense. It suggests that there is good crime-

---

[8] For a more detailed discussion, *see id.* at chs. 10 and 11.

[9] *See* MODEL PENAL CODE § 1.02 (Official Draft 1962); ARIZ. REV. STAT. ANN. § 13-101 (2009).

[10] MODEL PENAL CODE §1.02 (as amended May 16, 2007).

control utility in doing justice, and in that sense rests upon an instrumentalist rather than a deontological perspective. However, given the practical realities of assessing desert principles, it may be that empirical desert offers the best practical approximation of deontological desert, and for that reason the position may be highly attractive to the retributivist perspective.[11]

One final note is in order. Our focus here has been on how to construct the best principle for distributing criminal liability and punishment. Obviously, this is an important inquiry – the most important, the most fundamental in criminal law theory. However, we must not lose sight of the fact that in the larger scheme of things the issue of the proper distribution of criminal liability and punishment is only one of many important mechanisms for controlling crime, and perhaps not even the most significant.

Within the criminal justice system, the selection of policing and correctional policies can be of enormous importance. Outside of the criminal justice system, investments in education, social institutions, job training, mental health, and a host of other projects can have as or more important effects on the crime rate.

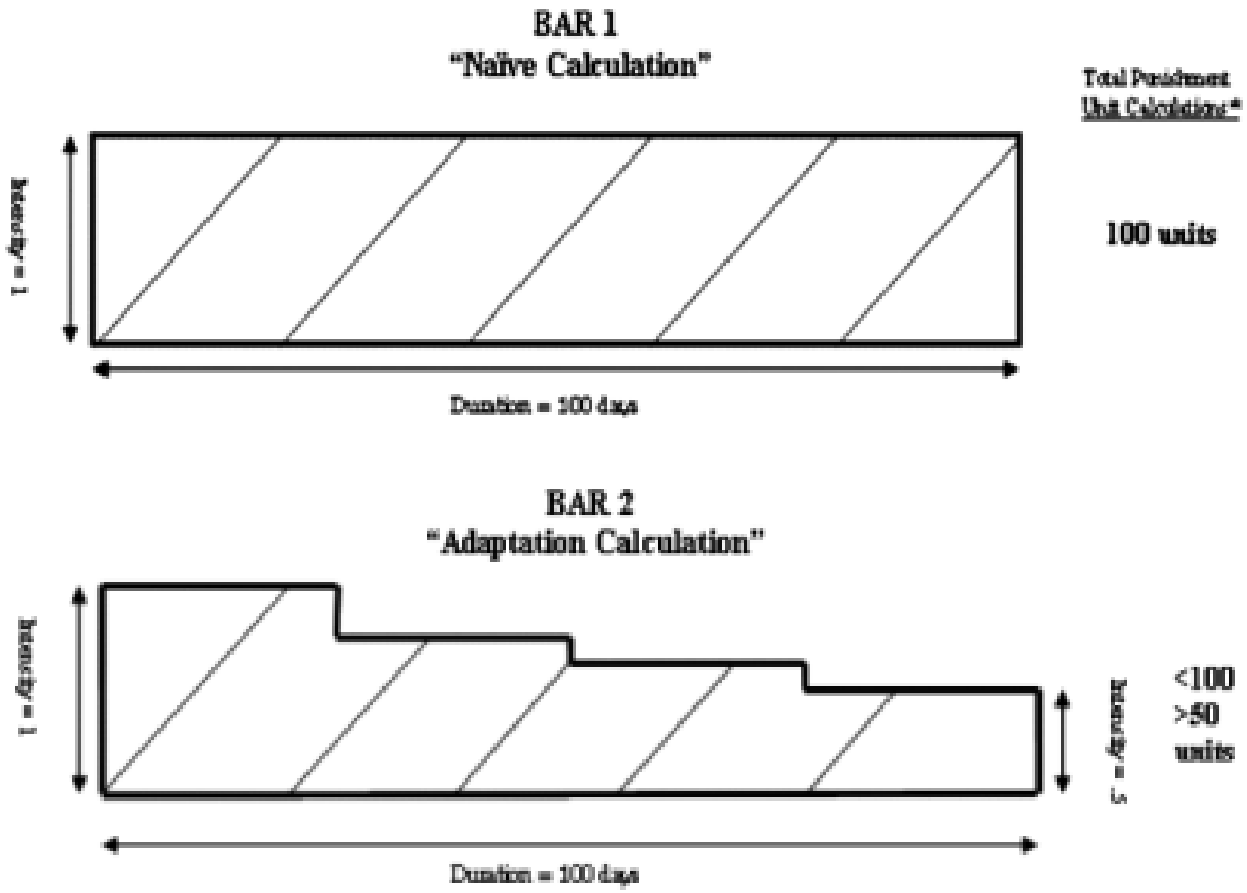
The point is this: don't expect too much from the criminal law's distributive principle for criminal liability and punishment. It is essential for doing justice and can make a significant contribution to crime control, but it can do only part of the work.

Thank you.

---

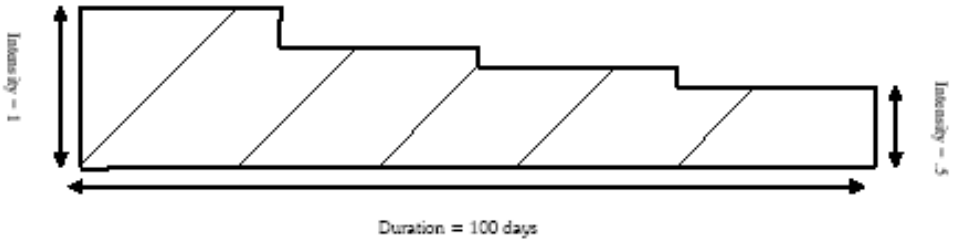
[11] *See* Paul H. Robinson, *The Role of Moral Philosophers in the Competition Between Philosophical and Empirical Desert*, 48 WM. & MARY L. REV 1831 (2007).

Distributive Principle	Distributive Criterion	Examples of Factors Relied Upon to Determine Amount of Punishment to be Imposed
Desert – deontological & empirical	Offender's blameworthiness	Nature & extent of harm, culpability level, justifying circumstances, personal capacities
Incapacitation	Offender's dangerousness	Period during which the offender remains sufficiently dangerous to need restraint
Rehabilitation	Offender's susceptibility to rehabilitation	Period required to rehabilitate the offender
Deterrence – special	What will deter crime by this offender	Proportional to seriousness of offense; more than that imposed for previous offense; perhaps low punishment for first offense
Deterrence – general	What will deter crime by others	Proportional to seriousness of offense; increase as publicity level increases; increase as capture rate decreases



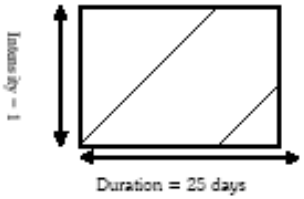
* Punishment unit calculation = Intensity x Duration = total area within the bar

BAR 2
"Standard Duration Calculation"



Standard "Duration" Calculation*	"Duration Neglect" Calculation**
<100 >50	.75+

BAR 3
"Duration Neglect Calculation"



25	1.0

* Standard "duration"
calculation: Intensity x Duration (total area within the bar)

** "Duration Neglect"
calculation: Maximum Intensity + End Intensity (memory of a longer duration is a minor extra feature, represented by a "+")

DP	Strengths/Advantages	Weaknesses/Disadvantages
GD	• under right conditions can avoid crime • can have effect beyond offender at hand	• works only when prerequisite conditions exist, which may not be common: knowledge of the deterrence-based rule, capacity and inclination to rationally calculate what is in one's best interest, perception that costs of crime (prob. x punishment) exceed benefits • can give punishment other than what is deserved
SD	• under right conditions can avoid crime by offender at hand	• works only on the offender at hand • works only when prerequisite conditions exist • can give punishment other than what is deserved

DP	Strengths/Advantages	Weaknesses/Disadvantages
RH	• if successful, can avoid crime by offender at hand • may have value in itself, in making person's life better	• works only on the offender at hand • only modest success in only limited cases • can give punishment other than what is deserved • problematic as sole DP; must combine with another • forcibly changing person's nature may raise ethical questions about intruding upon personal autonomy
Inc	• no doubt that effective in reducing crime by offender at hand	• works only on the offender at hand • commonly inaccurate in predictions, which causes wasted costs and unjustified detentions • can give punishment other than what is deserved • can better reduce prevention costs, reduce damage to system's moral credibility, and increase accuracy by operating as open civil preventive detention system apart from CJ system

DP	Strengths/Advantages	Weaknesses/Disadvantages
ED	• most likely to be seen as just punishment, which can increase criminal law's moral credibility, thus crime control benefit	• may do injustice that is not apparent to the present community • failure to deviate from empirical desert misses special crime control opportunities that can arise (e.g., deterrence prerequisites satisfied and high deterrent effect possible, reliable prediction of high probability of an offender's serious future offense)
DD	• does justice	• fails to prevent avoidable crime • difficulty in operationalizing because of common disagreement among moral philosophers

DP	Strengths/Advantages	Weaknesses/Disadvantages
GD	• under right conditions can avoid crime • can have effect beyond offender at hand	• works only when prerequisite conditions exist, which may not be common: knowledge of deterrence-based rule, capacity and inclination to rationally calculate what is in one's best interest, perception that costs of crime exceed benefits • can give punishment other than what is deserved
SD	• under right conditions can avoid crime by offender at hand	• works only on the offender at hand • works only when prerequisite conditions exist • can give punishment other than what is deserved
RH	• if successful, can avoid crime by offender at hand • may have value in itself, in making person's life better	• works only on the offender at hand • only modest success in only limited cases • can give punishment other than what is deserved • problematic as sole DP; must combine with another • forcibly changing person's nature may raise ethical questions about intruding upon personal autonomy
Inc	• no doubt effective at reducing crime by offender at hand	• works only on the offender at hand • commonly inaccurate in predictions, which causes wasted costs & unjustified detentions • can give punishment other than what is deserved • can better reduce prevention costs, reduce damage to system's moral credibility, and increase accuracy by operating as open civil preventive detention system
ED	• seen as most just, thereby optimizing criminal law's normative influence	• may do injustice that is not apparent to the present community • failure to deviate from empirical desert can miss special crime control opportunities (e.g., deterrence prerequisites present and high deterrent effect possible, reliable prediction of high probability of an offender's serious future offense)
DD	• does justice	• fails to prevent avoidable crime • difficulty in operationalizing because of common disagreement among moral philosophers