I. Expected Punishment: Punishment and Probability of Punishment

For the most part, criminal law is used to enforce property rules, not liability rules. The idea, in principle, is to prevent certain types of infractions from occurring at all. But in truth, we don't actually prevent every violation. Instead, the best we can do is create some probability of being caught. Why is this true? Why don't we try to raise the probability of punishment to 100%? The answer, of course, is increasing opportunity cost. Each increment of probability costs more and more, in terms of expenditures on law enforcement, innocent persons wrongly convicted, etc.

From the criminal's perspective, the relevant issue is expected punishment: the size of the punishment multiplied by the probability of it being imposed. For simplicity, we'll talk about punishments in dollar terms, although the punishments are not always fines. (For example, we might say a month in jail has a monetary value of $10,000 to some criminal.) In short,

\[ EP = qP \]

where P is the cost of a punishment to the criminal, and q is the probability of it being imposed. This expected punishment is what really affects the criminal's choice of whether to commit a crime. He will compare the expected punishment to the expected benefit of his crime, and commit the crime if and only if the benefit exceeds the expected punishment. (This is slightly simplified, because the criminal is not necessarily risk neutral. If he's risk averse, then the true cost to him is greater than the expected punishment.)

Now, let's suppose that we've decided we want to impose an expected punishment of \( EP' \). (How did we decide that? We'll discuss that in the next section. Take it as given for now.) Notice that there are multiple ways of delivering the same expected punishment. For instance, suppose \( EP' = 2000 \). One way to impose this expected punishment is with a 20% chance of a $10,000 fine (or the equivalent in jail time). Another way is with a 10% chance of a $20,000 fine. Another way is with a 5% chance of a $40,000 fine. Another way is with a 40% chance of a $5000 fine. Etc. We could represent this with a curve on a graph with probability on the horizontal axis, punishment on the vertical axis; it looks like an indifference curve (or more accurately, an isoquant).

So, which combination of probability and punishment should we pick? It might seem like we should pick a probability as small as possible. By decreasing the probability, we decrease the number of criminals we actually have to catch and prosecute. (This is true independent of any reduced deterrence effect, which we're ruling out by raising the punishment to keep a constant EP. The point is that with a smaller probability, we have to catch a smaller percentage of any given number of criminals.) We can save resources by firing cops, judges, prosecutors, etc.
The problem is that while probability is costly, so is punishment. The greater is the punishment per criminal, the more likely it is that the punishment will have to be in non-monetary form. $P = 5000$ could possibly be imposed as a fine, but higher $P$ might require jail terms. A fine is a nice form of punishment, because the criminal's loss is the rest of society's gain. But jail terms and other non-fine punishments are costly; the criminal loses, and so do the taxpayers. Consequently, there is an increasing cost to having higher $P$, just as there's an increasing cost to having higher $q$.

As a result, we are unlikely to pick a "corner solution" with infinitely high $P$ and infinitely low $q$ or vice versa. Different means of achieving the same $EP$ have different levels of social cost associated with them. We want to pick the combination of punishment and probability that achieves the given level of $EP$ at the lowest cost. This process should be familiar: It is nearly identical to the cost-minimization that a firm uses in the long-run to decide its optimal combination of labor and capital.

II. Optimal Deterrence

Using the cost-minimization process above, we can find out the minimum cost of achieving any given level of expected punishment ($EP$). But what level of $EP$ should we choose? This is closely related to the question of how much deterrence we want to achieve, or optimal deterrence.

As $EP$ rises, the number of crimes should decrease. So should we increase $EP$ until the number of crimes equals zero? No, because as noted earlier, the law of increasing opportunity cost (or diminishing returns) applies here. To increase $EP$, we have to increase both $q$ and $P$, and there are costs associated with both. Equal-sized increases in $EP$ will produce smaller and smaller reductions in crime; or, to say the same thing in a different way, equal-sized reductions in crime will require larger and larger increases in $EP$ to achieve.

To decide how much deterrence we want, we will use the usual marginal benefit/marginal cost approach. First, let's consider the marginal cost of deterring more crimes.

In general, we expect the marginal cost of crime prevention -- or anything else for that matter -- to be positive. But sometimes, it may turn out that the marginal cost is actually negative. How is this possible? Let's say we have an increase in the expected punishment. As a result, the number of crimes decreases (the deterrent effect). Consequently, there will be two offsetting effects on our total expenditure:

- an increase in the number and size of punishments, because we're catching a larger percentage of the offenders and punishing them more severely; and
- a decrease in the number of punishments, because the total number of offenders is smaller than before.

Which of these effects predominates? It totally depends on the numbers. If the behavior of criminals is relatively elastic with respect to $EP$-- so that a small increase in $EP$ leads to a large reduction in crime -- then the second effect will predominate, and total...
expenditure will fall. If the behavior of criminals is relatively inelastic with respect to EP, on the other hand, total expenditure will rise.

But eventually, the marginal cost will become positive and keep on rising. Why? For the same reasons we discussed earlier. Deterrence involves raising EP, which means raising the punishment and/or the probability of punishment -- probably a combination of the two. And as they get larger and larger, they will be increasingly expensive. The punishments will shift more and more toward those costliest to society (like jail time), and the probability of punishment is subject to the law of increasing opportunity cost (including the possibility of punishing innocents). [Draw upward-sloping MC curve, with the left end below the horizontal axis.]

Thus, we have an upward-sloping marginal cost of deterrence. Now let's consider the marginal benefits. The benefit of deterrence, from an efficiency perspective, is preventing value-decreasing transactions. For instance, if a thief steals your TV, and he values the TV less than you do, then the theft decreased total social wealth. So we'd like to prevent the theft. But how much are we willing to pay to prevent it? Let's say the TV is worth $400 to you and $300 to the thief. Then the theft decreased social wealth by $100. Now suppose it would cost $200 to prevent this theft (with no additional deterrent effect on other thieves). Then it would not make sense to try to deter this theft, because it costs more than it's worth. (Note the similarity to the discussion of commons and privatization. Commons may be inefficient, yet not inefficient enough to justify privatization if the costs of enforcing private property are too high.)

We said earlier that a potential criminal will compare the expected punishment to his private benefit from the crime. In other words, he will commit the crime if and only if

\[
\text{private benefit to criminal} > \text{EP}
\]

In the example above, the criminal will steal the TV if the expected punishment is less than $300. So, what could we do to prevent this theft? We could increase the EP so it's just barely above $300 -- say, $301. By doing so, we'd prevent a loss of wealth equal to $100. So the marginal benefit of deterring this crime is $100.

Now suppose there's another thief who values his neighbor's television at $325 (and his neighbor values it at $400, as before). This thief will not be deterred by our new, higher punishment of $301. To deter him, we'd have to raise the punishment to $326. By doing so, we'd prevent a loss of wealth equal to $75. So the marginal benefit of deterring this crime is $75, which is positive but less than the benefit of deterring the last crime. Following this logic, we observe that marginal benefit is likely to be downward-sloping. Why? Because the first crimes deterred are those worth least to the criminals, while the last crimes deterred are those worth most to the criminals.

It's even possible that MB will eventually become negative. Why? Because large enough punishments will deter some efficient crimes. An efficient crime is one that creates a greater gain to the criminal than loss to the victim. In many cases, we need not
worry about such efficient crimes, because the same transaction can be arranged voluntarily; e.g., if the thief values your TV at $450 while you value it at $400, he can just buy it from you. But sometimes a voluntary transaction is not possible; consider DDF's example of a starving hunter who breaks into a private cabin and eats the food. Because the owner is not present, a voluntary transaction is not possible, but the benefits to the hunter almost certainly exceed the loss to the owner. Once we start deterring "crimes" like these, the MB of deterrence can become negative. [Draw downward-sloping MB curve, with the right end below the horizontal axis.]

So now we can put our MC and MB together in order to find the optimal amount of deterrence. The optimal number of crimes deterred occurs where they cross. This means that the marginal benefit of the deterrence is greater than the marginal cost of the last crime deterred.

### III. Optimal Punishment

For this optimal number of crimes deterred, there will be a corresponding EP, which we'll implement using some combination of punishment and probability of punishment. But we can be a little more specific.

At the optimal level of deterrence,

\[
MC \text{ of deterrence} = MB \text{ of deterrence}
\]

And we know that the MB of deterrence is the difference between the valuation of the victim and the criminal: \( MB \text{ of deterrence} = \text{loss to victim} - \text{gain to criminal} \). So:

\[
MC \text{ of deterrence} = \text{loss to victim} - \text{gain to criminal}
\]

We discovered above that, on the margin, a criminal will be deterred when the EP is raised to just above the private benefit of the crime to him. The thief's private benefit from the TV was $300, and it took a $301 expected punishment to prevent the theft. So at the margin, the expected punishment equals the gain to the criminal (i.e., whatever is the current EP, it's equal to the gain to the criminal of the last crime deterred). Thus:

\[
MC \text{ of deterrence} = \text{loss to victim} - \text{expected punishment}
\]

Rearranging this, we find:

\[
\text{expected punishment} = \text{loss to victim} - \text{MC of deterrence}
\]

How can we interpret this?

- First, imagine that \( MC = 0 \). That is, we can prevent crimes at zero cost. If so, then we would set the expected punishment equal to the loss to the victim. Then the only crimes that would occur would be those where the gain to the criminal is greater than
the damage to the victim. But crimes like those are actually efficient (value increasing).

• Second, imagine that $MC > 0$. That is, it's costly to prevent crimes. This is most likely the case when criminal behavior is inelastic with respect to $EP$. In that case, we would set the expected punishment below the loss to the victim. The result is that some crimes will take place that are value-decreasing; these crimes will involve criminals who gain more than the $EP$ but less than the loss to the victim. Why don't we deter these crimes? Because it's too costly to do so; the gains (in terms of preventing a value-reducing crime) aren't enough to justify the cost.

• Third, imagine $MC < 0$. That is, the legal system actually saves money by preventing crimes; see above for explanation of how this could be. In this case, we would set the expected punishment above the loss to the victim. The result is that some efficient crimes will be deterred. Why do we deter them? Because doing so allows us to avoid the inefficient crimes at lower cost. [Show that this would occur when $MC$ and $MB$ are both negative, crossing below the horizontal axis.]

We've seen this argument before, in the discussion of why there are punitive damages in tort law. The argument is that some torts are easier to deter, while others are more difficult to deter. For the easier to deter torts, the $MC$ may be negative, in which case it makes sense to have an expected punishment ($EP$) greater than the loss to the victim -- which is what punitive damages do. For more difficult to deter torts, $MC$ is probably positive, in which case it makes sense to have an $EP$ less than the loss to the victim -- which is exactly what happens with a less-than-100% chance of being held liable and no punitive damages.

Another application: Should first degree and third degree murder be punished equally? After all, a murder is a murder -- the loss to the victim is pretty much the same. But a planned murder is probably easier to deter than a crime of passion, because there's more time for consideration. So the marginal cost of deterring third degree murder is lower, meaning the expected punishment should be higher.

IV. **Should the Rich and the Poor Get Equal Punishments?**

DDF considers the question of whether the rich should get higher punishments than the poor. Many people's intuitive answer is yes, but that's not necessarily true (although it could be).

Suppose that the $MC$ of deterring offenses is zero, for both rich and poor. Then the expected punishment should be set equal to the loss to the victim; that way, all and only inefficient crimes are deterred. This is true for both rich and poor. Is it possible that the rich will commit more offenses as a result? Maybe, if for some reason their gain from committing offenses is greater than the gain to a poor person. But that's fine -- they're still committing only efficient offenses. They may buy more of them, but they buy more CD's, too.
Now suppose that the MC of deterring offenses is positive, but still the same for rich and poor. Then expected punishment should be less than the loss to the victim; that way, inefficient offenses will be deterred if and only if the gain from deterring them is less than the cost. And again, the rule should be the same for both rich and poor.

The story changes if MC differs between rich and poor. Suppose the MC is greater for the rich than for the poor. Why might this be true? (a) For some offenses, the dollar value to the rich is greater than the dollar value to the poor, because the marginal utility of money is lower for the rich. So it will take a larger expected monetary punishment to get the same amount of deterrence. (b) The rich can hire better lawyers, so a greater expenditure is necessary to get the same expected punishment. Either way, it's more expensive to deter rich criminals. Using the formula above, the implication is the EP should be lower for the rich.

Now suppose the MC is less for the rich than for the poor. Why? (a) Because the same dollar-valued EP is less expensive to impose on the rich, since it can be done with fines instead of jail time. (b) For offenses that involve jailtime, the opportunity cost of a given amount of time is greater for the rich than for the poor, so a smaller amount of jailtime could have the same deterrent effect. Either way, the result is a lower MC for the rich. Using the formula above, the implication is the EP should be higher for the rich.

V. Why Consider Gains to Criminals?

In the analysis above, we considered the gains to criminals as part of the calculus. In finding out the marginal benefit, we were really finding a net marginal benefit, because we were looking at the net loss in wealth after subtracting the gains to the criminal. Why are we doing this?

If you already take it as given that certain things are intrinsically, morally wrong, then you might measure the benefits of law enforcement differently. But we don't want to assume what we're trying to prove. Why do we have the moral beliefs that we have? The efficiency approach to law suggests that, in part, our moral beliefs may result from implicit efficiency considerations. Maybe the reason so many societies have morally opposed theft is that theft tends, on the whole, to be value-decreasing. It transfers goods from people who value them more to people who value them less. It also decreases the incentive to be productive, since theft is like a kind of tax on your effort. A society that did not have a taboo against theft would probably be an impoverished society. So the efficiency approach to law says to set aside, at least for the time being, other value considerations and look only at wealth. If it turns out that some of our moral beliefs derive from efficiency considerations, it would be double-counting if we included those moral considerations as part of the benefit of law enforcement -- because we already included the wealth considerations underlying those beliefs.

But suppose you believe that some moral considerations go beyond efficiency. That's a perfectly respectable position (although one that deserves careful thought, not just assertion). Does that mean the above analysis is inapplicable? No. Even if you don't
think efficiency is the only consideration, you can include whatever other (moral) benefits of law enforcement that you want in the MB curve.

Say, for instance, that you oppose prostitution or drug use on moral grounds; these are voluntary (hence value-increasing) transactions that you want to prevent. Nonetheless, you wouldn't be willing to spend an infinite amount of money preventing either activity; diminishing returns demand that you stop at some point, or else sacrifice all your other values to this one. There will be an increasing marginal cost of deterring these activities, and presumably a decreasing marginal benefit (albeit measured differently from before). And these considerations indicate, once again, that the MC = MB rule is the appropriate way to balance your values.

VI. A Few Caveats

There are few things we have to keep in mind for our analysis to be accurate:

- **Direct deterrence.** We've been talking about indirect deterrence. But direct deterrence could also be an argument for having larger punishments on individuals we've already caught: we prevent this particular person from committing the same crime again. These benefits would have to be included in MB.
- **Rehabilitation.** If a punishment actually makes the criminal less likely to commit crimes in the future (or otherwise makes him a more valuable citizen), that should be included as an increase in the MB. On the other hand, if punishment actually makes the criminal more harmful (he becomes a more hardened criminal by exposure to others), that should be included as a reduction in MB.
- **Marginal deterrence.** We should take into account that criminals may substitute one crime for another. If we increase the punishment of one crime, we may induce the criminal to switch to a different kind of crime. For instance, punishing prostitution may induce some prostitutes to sell drugs instead.

VII. Less Expensive Punishments

It was argued earlier that the MC of deterring crimes is generally positive and increasing because as the severity of the punishment rises, it's more likely that executing the punishment is costly to society as well as to the criminal. A fine is a benefit to society as well as a cost to the criminal, but a prison term is a cost to everyone.

However, there are some forms of punishment that appear not to have this feature, or at least to have it in less severe form. What about, say, the death penalty? There is a one-time cost of imposing it, but it avoids the cost of keeping someone in prison. (In the status quo, it costs a lot to have someone on death row, but that's mainly a function of public opposition and the legal regime, not the penalty itself.) Plus, we can impose this higher punishment with lower probability, thereby reducing the number of offenders we must punish. Taken to its logical conclusion, we could have a regime in which all offenses are punished with death, with a very low probability of punishment. Think, for example, of Saudi Arabia, where the crime rate is very low.
There are other punishments that may even be a benefit to society (negative marginal cost). These include chain gangs, prison factories, etc. And to reduce the cost of the death penalty, the criminals' organs could be forfeited and sold at auction. Or their cadavers could be sold to medical schools. These types of punishment would reduce the marginal cost of punishment dramatically, thus allowing a much higher amount of deterrence. Why don't we do these things?

There are some obvious answers, but as DDF notes, they are not quite adequate. (1) The fact that execution (and many other punishments, such as removal of body parts) is irreversible. This is true, but then again, years spent in prison are also irreversible. The remainder of a prison term can be commuted in the case of mistakes, but how often does that really happen? And besides, if there's a greater deterrent effect, there will be fewer mistakes to correct. Fewer murders implies fewer trials and hence fewer wrongful convictions. (2) Execution is actually expensive in the status quo, as noted above. But this is not a necessary consequence of the punishment itself. We could impose death with exactly the same safeguards as any other punishment. (3) We are morally opposed to killing people. This is actually the real reason that execution is currently expensive. But we can't just take this as given; the question is why we are so opposed (and whether we should be).

DDF proposes the "public choice" explanation: we have to consider the incentives such punishments give to the agents of the state. If political actors can benefit from imposing punishments (by providing public services with lower taxes, for example), this provides them with an incentive to capture more people within the legal system, for even the most minor of offenses. This is, in essence, a form of rent-seeking. It is also sometimes referred to as “net widening.” An examples of such net widening in the status quo is the civil asset forfeiture laws, which allow police departments to take the assets of suspects and not return them (even if the suspect is found innocent). Naturally, this gives the police a huge incentive to charge more suspects, and that's exactly what they've done.

Although I think the argument just given is a good explanation for why punishments that provide benefits to the state (such as organ forfeiture) are undesirable from an efficiency perspective, I don’t think the argument works as well for the death penalty. Although execution could be low cost, it does not clearly provide transfers of wealth to agents of the state.

VIII. The Crime/Tort Distinction

Would it be possible to convert all criminal law into civil law? Historically, it began that way in England. Criminal law did not emerge until the king began "nationalizing" certain offenses as a means of raising money. He classified certain offenses as "crimes against the crown," and he created a group of judges who would "ride circuit" to hear such cases on appeal -- for a price. Such was the beginning of the criminal law system.
How would a fully civil system work? What we currently know as crimes would be classified as torts instead. People who were assaulted or raped would sue for compensation; relatives of people killed would do the same. Intuitively, this seems bizarre and unworkable; but what are the objections?

1. "The victim of an offense may not have sufficient resources to prosecute it." This is a problem that could be dealt with through transferable tort claims, meaning a victim could sell the right to prosecute the case to someone else. Alternately, plaintiffs could raise money from the equivalent of venture capitalists (which is really just a means of selling part of one's expected damage awards). True, the victim might end up with less than total compensation for his loss, but that's true in the status quo as well: the criminal justice system gives him nothing at all.

2. "Some offenses cause diffuse injury, so nobody has an adequate incentive to prosecute them." The status quo civil system deals with this problem through class action lawsuits. DDF suggests it could be dealt with even more easily if tort claims were transferable.

3. "Some offenses result in a diffuse injury difficult to observe, along with an observable injury to a single victim." For example, some crimes may hurt one person while acting as an implicit threat to others. This could be dealt with through punitive damages. (We talked about this before when we discussed “strategic torts.”)

4. "If an offender is judgment-proof, there is no incentive for the victim to prosecute him, so prosecution must be by the state." One solution to this problem is to allow penal slavery, collection of organs, and other punishments that actually provide benefits to the punisher. (Such punishments would suffer from the same problem discussed earlier: they could lead to rent-seeking.) Or, the state could provide a "bounty" system that would pay extra benefits to the private prosecutors in the case of judgment-proof defendants. But although this would give sufficient incentives to prosecute crimes that have already occurred (because the victim/prosecutor gets fully compensated), it would not give an adequate incentive to offenders not to commit crimes in the first place (because part of the compensation comes from the state, not the offender). So the state might have to step in with criminal punishments in such cases, while also paying bounties to the private prosecutors. Effectively, this system would maintain a criminal system while privatizing prosecution.

5. “Deterrence of criminal acts is a public good. (Or, deterrence of criminal acts is a positive externality that results from prosecution.)” In other words, when you prosecute someone, you provide the benefit of sending a message that crimes will be prosecuted. Everyone benefits from this message having been sent; the benefit is non-excludable. But is this really true? Only for some crimes. If you hire a private agency to prosecute criminals who break into your house, that doesn’t send a message about any house other than yours and others protected by the same agency. The public good argument only works under more narrow circumstances: (1) The crime has anonymous victims; that is, the criminal can’t tell when he commits the crime
whether his selected victim has paid for protective services. As a result, anyone who pays for protective service raises the overall likelihood that a randomly chosen victim has protection. The benefits are dispersed, while the cost is concentrated. [Compare John Lott’s argument for concealed-carry laws.] (2) Direct deterrence -- i.e., containing or quarantining the criminal -- is a major component of overall deterrence. In that case, punishing the criminal by putting him away provides a service to everyone. Both of these types of situation could, possibly, be dealt with by providing bounties, as discussed in 4 above.

6. "It is impossible to construct legal rules that give victims the optimal incentive to prosecute." There is an optimal combination of punishment and probability of punishment. But even if the state controls the amount of punishment, privatizing prosecution takes the probability out of its hands. There is no guarantee that the right number of private prosecutions will take place. To put it another way, under the status quo, the state has control over two variables: punishment and probability. If prosecution were privatized, it would have control only over punishment, not probability.

There is a more streamlined way to think about all of these issues together: Ask whether the right to prosecute and punish a criminal can be sold at a positive or negative price. (Positive price: someone pays me money for the right to prosecute and punish an offense against me. Negative price: I pay someone else to prosecute and punish an offense against me.) Now, what determines whether the right has a positive or negative price?

- The right has a positive price if the expected gain to the prosecutor is greater than the marginal cost of catching and punishing the offender. In such a case, a private prosecution firm would be willing to pay to acquire the right to prosecute from the victim.
- The right has a negative price if the expected gain to the prosecutor is less than the marginal cost of catching and punishing the offender. In such a case, a victim would have to pay a prosecutor to prosecute. The prosecution is a service provided by the prosecution firm to the victim. Why would a victim pay to have this done? In order to send a message to future offenders: If you hurt me, I will fight to have you punished.

Now, we can expect that any offense for which the right to prosecute and punish has a positive price would be prosecuted in a fully private system, either by the victim or by someone who purchased the right to prosecute from the victim. We can also expect that any offense for which the right to prosecution and punishment has a negative price will be prosecuted in a fully private system -- if deterrence is a private good for the offense in question. So the Achilles’ heel of privatized prosecution, it would seem, is offenses that are not private goods and for which the right to prosecute and punish has a negative price. These will mostly correspond to anonymous victim offenses.

If the current legal system is efficient, then the sorting of offenses between civil and criminal should mirror the typology above. Does it?
Here’s one piece of evidence in favor: When offenders are likely to be judgment-proof, the prosecution right is likely to have a negative price. A private prosecution firm would not buy the right to prosecute from you; you’d have to pay them to take on the responsibility. This is especially true if the prosecutor is also responsible for paying for any positive-cost punishments such as imprisonment. Offenses with judgment-proof defendants are, indeed, often treated as criminal rather than civil in the status quo.

Evidence against: Under the status quo, the torts we actually see getting prosecuted are those with positive price. Otherwise, lawyers would not find it worthwhile to prosecute them. So we are observing a biased sample, because there could be many other torts that simply don’t get brought to court.

More evidence against: Just because an offense has a negative price, that doesn’t mean it’s also an anonymous victim offense. And in fact, the offenses we currently treat as crimes are typically not anonymous victim offenses: rape, murder, burglary, etc. Many offenses that we treat as torts are anonymous victim offenses: car accidents, for example.

For these and other reasons, it is difficult to make the case that offenses are sorted efficiently between civil and criminal law.