

Journal of Legal Studies,
forthcoming January 1999.

**On the Disutility and Discounting of Imprisonment
and the Theory of Deterrence**

A. Mitchell Polinsky

and

Steven Shavell*

Abstract: This article studies the implications for the theory of deterrence of (a) the manner in which individuals' disutility from imprisonment varies with the length of the imprisonment term; and (b) discounting of the future disutility and future public costs of imprisonment. Two questions are addressed: Is deterrence enhanced more by increasing the length of imprisonment terms or instead by

* Stanford Law School and Harvard Law School, respectively; both authors also are research associates of the National Bureau of Economic Research. Polinsky is grateful for research support from the John M. Olin Program in Law and Economics at Stanford Law School, and Shavell acknowledges research support from the John M. Olin Center for Law, Economics, and Business at Harvard Law School. The authors also thank Steven Levitt for comments and Mark Liffman for research assistance.

raising the likelihood of imposing imprisonment? What is the optimal combination of the severity and probability of imprisonment sanctions?

1. Introduction

In this article, we show that the manner in which individuals' disutility from imprisonment varies with the length of the imprisonment term has important implications for the theory of deterrence. We distinguish among three possibilities: disutility from imprisonment rises in proportion to the length of the imprisonment term; disutility rises more than in proportion to the term (suppose that prison becomes increasingly difficult to tolerate); and disutility rises less than in proportion to the term (suppose that the initial period of imprisonment involves great disutility or that individuals become inured to prison life over time).¹ We also show that the discounting by individuals of the future disutility of imprisonment, as well the discounting by society of the future public costs of imprisonment, bear significantly on deterrence theory.

We first discuss the implications of the form of disutility and discounting for a time-honored positive question:² Is deterrence enhanced more by increasing the length of imprisonment terms or instead by raising the likelihood of imposing imprisonment? We observe that if the disutility from imprisonment rises in proportion to sentence length, then an increase in the magnitude of sanctions has the same effect as an equal percentage increase in the probability of sanctions. However, if disutility rises more than in proportion to the sentence, raising the magnitude of sanctions has a greater effect than increasing their probability by the same percentage amount. Conversely, if disutility rises less than in proportion to the sentence, raising the magnitude of sanctions has a smaller effect than increasing their probability. When individuals' discounting of disutility is taken into account, the conclusion is qualitatively the same as in the last case -- the magnitude tends to have a smaller effect than the probability.

¹ As will be discussed, each of these forms of disutility is associated with an attitude toward the risk of imprisonment.

² This question was raised over two hundred years ago. See Cesare Beccaria, *On Crimes and Punishments and Other Writings* Ch. 27 (edited by Richard Bellamy, 1995, originally published 1764).

We then consider a basic normative question: What is the optimal combination of the severity and probability of imprisonment sanctions? We demonstrate that maximal sanctions, applied with correspondingly low probabilities, are optimal when the disutility of imprisonment rises either proportionally or more than proportionally with the length of the sentence. In essence, this is because raising the sentence and lowering the probability so as to maintain deterrence saves enforcement costs and either does not affect the public cost of imprisonment (if the disutility of imprisonment rises proportionally) or causes it to decline (if disutility rises more than proportionally). But we also show that less-than-maximal imprisonment sanctions, applied with correspondingly high probabilities, may be optimal when the disutility of imprisonment rises less than proportionally with the sentence. This result follows because raising the sanction and lowering the probability so as to preserve deterrence now causes the public cost of imprisonment to rise.

A similar conclusion can occur when discounting is taken into account. In the case that we believe is most plausible -- when the rate at which individuals discount the future disutility of imprisonment exceeds the rate at which society discounts the future costs of imprisonment -- less-than-maximal sanctions may be optimal. This is because raising the sanction and lowering the probability so as to maintain deterrence causes the discounted public cost of imprisonment to rise. However, if the individual discount rate is less than (or equal to) the social discount rate, maximal sanctions are optimal for reasons analogous to those discussed above.

Section 2 describes our assumptions about disutility and discounting in more detail. Section 3 discusses the positive question concerning the relative effectiveness of raising the magnitude of the sanction versus the probability to increase deterrence. Section 4 addresses the normative question of the optimal magnitude and probability of imprisonment sanctions. Section 5 discusses the applicability of our results and their relationship to the corresponding analysis of monetary sanctions. An Appendix contains proofs of the results in Section 4.³

³ The contribution of this article is twofold. First, we analyze the implications of different assumptions about the disutility of imprisonment more systematically than has previously been done. Second, we formally incorporate the factor of discounting (both private and social) into the analysis of imprisonment sanctions for the first time. For further discussion of the relationship between our article and prior literature, see notes 18 and 29 below.

2. The Disutility and Discounting of Imprisonment Sentences

We assume that each year a person is in prison, he experiences disutility. One possibility is that the disutility is equal to a constant, c , from one year to the next. In this case, the total disutility, $d(s)$, from s years of imprisonment is sc ; total disutility rises proportionally with the length of the sentence.

Another possibility is that the disutility suffered by a person each year increases with the sentence length. This form of disutility would describe a person for whom prison becomes increasingly difficult to tolerate as time passes, or for whom separation from family, friends, and everyday life becomes more and more painful. For such a person, the total disutility of imprisonment $d(s)$ rises more than in proportion to the sentence length.

A different possibility is that the disutility experienced each year declines over successive years of imprisonment. Disutility of this form might arise because a person becomes accustomed to prison life or because he ceases to care as much about those he knew from the outside. Also, the disutility associated with the first year of prison might be particularly great compared to that of later years for the following reasons: humiliation and brutalization of the prisoner may occur early on and do its major harm then; and stigmatization of the prisoner (which lowers earning capacity and status) may be due primarily to having been in prison at all, and not increase much with the number of years spent there. If the disutility per year falls for each year of imprisonment, or only from the first year to a constant level thereafter, total disutility $d(s)$ rises less than in proportion to the sentence length.⁴

The disutility that a person experiences from imprisonment in any future year might be discounted to the present at some positive rate. Discounting is a potentially significant factor, especially because many criminologists believe that criminals tend to be more present-oriented than the general population.⁵ If the immediate disutility each year is a constant c and is discounted at a rate r , the present value of the stream of disutility for a sentence of s years is

⁴ To see this when the disutility from the first year is higher than an annual constant thereafter, let c_1 be the disutility from the first year and c be that from each succeeding year, where $c_1 > c$. Then the average disutility for s years is $[c_1 + (s - 1)c]/s = (1/s)c_1 + [(s - 1)/s]c$, which is decreasing in s .

⁵ See, for example, James Q. Wilson and Richard J. Herrnstein, *Crime and Human Nature* 416-21 (1985) and John J. DiIulio, *Help Wanted: Economists, Crime and Public Policy*, 10 *J Economic Perspectives* 3, 16-17 (1996). Of course, one might expect that those individuals in the population who are most present-oriented would

given by the discounted total $d(s) = c + c/(1+r) + c/(1+r)^2 + \dots + c/(1+r)^{s-1}$. In this case, the total discounted disutility from imprisonment increases less than in proportion to sentence length;⁶ for example, if an imprisonment term of 5 years is doubled to 10 years, the total discounted disutility increases by less than a factor of two (for example, by only 62% if the discount rate is 10%). Thus, the discounting of a constant stream of disutility may be viewed as a special case of disutility that rises less than in proportion to sentence length.

In general, if c_i is the immediate disutility from the i th year of imprisonment, total discounted disutility is $d(s) = c_1 + c_2/(1+r) + c_3/(1+r)^2 + \dots + c_s/(1+r)^{s-1}$. Discounting can occur regardless of whether the immediate disutility experienced each period rises, falls, or remains constant. For expositional convenience, we will assume when we discuss discounting that the disutility per period is constant.

gravitate toward crime, since the discounted expected disutility of sanctions is lower for them than for others.

⁶ Specifically, since $d(s+1) - d(s) = c/(1+r)^s$ is decreasing in s , $d(s)$ increases less than in proportion to s .

3. The Effects of the Magnitude versus the Probability of Imprisonment

Sanctions on Deterrence

In this section we examine how deterrence is influenced by the magnitude and the probability of imprisonment sanctions. We assume that deterrence is determined by the *expected disutility* of sanctions: if the probability p of an imprisonment sentence, multiplied by its total disutility $d(s)$, exceeds the benefit a person would derive from a criminal act, he will be discouraged from committing it.

First consider the case in which total disutility rises proportionally with the length of the sentence -- that is, total disutility equals the sentence length multiplied by a constant, the amount of disutility per year. We will describe a person for whom this is true as *risk neutral* in imprisonment.⁷ For risk-neutral individuals, the severity and the probability of imprisonment sanctions have equivalent influence on deterrence, in the sense that a given percentage increase in either has the same effect on deterrence.⁸ To illustrate, suppose initially that p is .5, s is 10 years, and the constant disutility per year is 3, resulting in expected disutility of 15 ($= .5 \times 10 \times 3$). If p is raised by 20% to .6, then expected disutility rises to 18 ($= .6 \times 10 \times 3$). If instead s is raised by 20% to 12, then expected disutility also rises to 18 ($= .5 \times 12 \times 3$).⁹

Next consider total disutility that increases more than in proportion to the number of years of imprisonment.¹⁰ For example, suppose that the total disutility of imprisonment for a

⁷ This term is used because the person is indifferent between -- feels neutral about the choice between -- a certain sentence of, say, 10 years, and a risky sentence with the same expected value of 10 years, such as a sentence of 8 years imposed with a 50% chance and a sentence of 12 years imposed with the same chance ($(.5 \times 8) + (.5 \times 12) = 10$). If the disutility per year of prison is 3, then the certain 10-year sentence has disutility of 30, and the expected disutility of the risky sentence also is 30 ($(.5 \times 8 \times 3) + (.5 \times 12 \times 3) = 30$).

⁸ It is necessary to discuss the influence of the probability and of the magnitude of a sanction in terms of their *percentage* changes because, otherwise, the choice of units of measurement would make any comparison meaningless. For example, if imprisonment sanctions were measured in hours, then the effect of a one unit increase in the sanction would be small -- one hour more of imprisonment would hardly affect deterrence -- compared to a one unit -- say, one percent -- increase in the probability. If instead imprisonment were measured in years, a one unit increase in the sanction could be substantial compared to a one unit increase in the probability. This problem -- that comparisons that depend on the choice of units of measurement are arbitrary -- can be avoided by expressing the changes in the sanction and in the probability in percentage terms.

⁹ It is easy to see in general that the severity and probability of sanctions have equivalent effects on deterrence in the present case: If the sanction rises from s to $(1 + \Delta)s$, expected disutility rises from psc to $p(1 + \Delta)sc$, while if the probability rises from p to $(1 + \Delta)p$, expected disutility rises to $(1 + \Delta)psc$, the same thing. (Note that, for a risk-neutral person, expected disutility, and thus deterrence, depends only on the expected value of the length of the prison sentence, ps .)

¹⁰ Formally, the assumption is that $d'(s) > d(s)/s$.

person is the number of years he is in prison squared. We will describe such a person as being *risk averse* in imprisonment.¹¹ For risk-averse individuals, the severity of imprisonment sanctions has a greater effect on deterrence than the probability of such sanctions; that is, a given percentage increase in the length of the sentence raises deterrence more than does the same percentage increase in the probability of the sanction being imposed. If the disutility of imprisonment is the square of the number of years, if p is .5, and if s is 10 years, then expected disutility is 50 ($= .5 \times 10^2 = .5 \times 100$). Now if the sanction rises by 20% to 12 years, expected disutility increases to 72 ($= .5 \times 12^2 = .5 \times 144$). If instead the probability rises by 20% to .6, expected disutility increases only to 60 ($= .6 \times 10^2 = .6 \times 100$), so deterrence is not as great.¹²

Now suppose that total disutility grows less than in proportion to the length of the imprisonment term.¹³ For instance, total disutility might equal the square root of the number of years in prison, or disutility might be high the first year and equal to a smaller constant amount each subsequent year, such as 200 the first year and 10 each later year. A person whose disutility grows less than in proportion to sentence length may be called a *risk preferrer* in imprisonment.¹⁴

For risk-preferring individuals, the severity of imprisonment sanctions has a lesser effect on deterrence than the probability of sanctions; a given percentage increase in the length of the sentence raises deterrence less than does the same percentage increase in the probability of the sentence being imposed. To illustrate, if the disutility of the first year of imprisonment is 200 and the disutility each year thereafter is 10, if p is .5, and if s is 10 years, then expected disutility initially is 145 ($= .5[200 + (9 \times 10)]$). If s is raised by 20% to 12 years, expected disutility

¹¹ This term is employed because the person is averse to pure risk in his sentence. For instance, he would prefer a certain sentence of 10 years to a risky sentence with the same expected value, such as an 8 year sentence with a 50% chance and a 12 year sentence with a 50% chance. If the disutility of imprisonment equals the square of the sentence, the disutility of the certain sentence is 100 ($= 10^2$), whereas the expected disutility of the risky sentence is 104 ($= (.5 \times 8^2) + (.5 \times 12^2) = 32 + 72$), which is higher.

¹² More generally, we want to show that expected disutility is higher if s rises to $(1 + \Delta)s$ than if p rises to $(1 + \Delta)p$. Equivalently, we want to show that $pd((1 + \Delta)s) > (1 + \Delta)pd(s)$. But this follows from $d((1 + \Delta)s) > (1 + \Delta)d(s)$, which is true since $d'(s) > d(s)/s$.

¹³ Formally, the assumption is that $d'(s) < d(s)/s$.

¹⁴ This term is used because the person prefers risk in his sentence. He prefers a risky sentence of 8 years with a 50% chance and 12 years with a 50% chance to a certain sentence of 10 years. If his disutility equals the square root of the sentence length, then his expected disutility from the risky sentence is 3.15 ($= (.5 \times \sqrt{8}) + (.5 \times \sqrt{12}) = 1.414 + 1.732$), whereas his disutility from a certain 10-year sentence is 3.16 ($= \sqrt{10}$), so the risky sentence is preferred.

becomes 155 ($=.5[200 + (11 \times 10)]$), whereas if instead p is raised by 20% to .6, expected disutility becomes 174 ($=.6[200 + (9 \times 10)]$), which is higher.¹⁵

Finally, consider discounting when disutility per year in prison is constant. In this case, as noted above, total disutility rises less than in proportion to sentence length. Such a person is a risk preferrer. Thus, *if individuals discount the future disutility of imprisonment and disutility per year is constant, the magnitude of an imprisonment sanction has a lesser effect on deterrence than the probability of the sanction being imposed.* For example, suppose that disutility is 10 per year and is discounted at a rate of 25%, and that initially p is .5 and s is 10 years. It can be calculated that the present value of the disutility of the 10-year sentence is 44.63,¹⁶ so the expected disutility is 22.32 ($=.5 \times 44.63$). If s is raised by 20% to 12 years, the expected disutility becomes 23.28 ($=.5 \times 46.56$). But if instead p is raised by 20% to .6, the expected disutility becomes 26.78 ($=.6 \times 44.63$), which is higher.^{17,18}

¹⁵ In general, we want to show that expected disutility is lower if s rises to $(1 + \Delta)s$ than if p rises to $(1 + \Delta)p$. Equivalently, we want to show that $pd((1 + \Delta)s) < (1 + \Delta)pd(s)$. But this follows from $d((1 + \Delta)s) < (1 + \Delta)d(s)$, which is true since $d'(s) < d(s)/s$.

¹⁶ The numbers in the text may be determined from the formula $c + c/(1 + r) + \dots + c/(1 + r)^{s-1} = c[1 - (1/(1 + r)^s)]/[1 - (1/(1 + r))]$. For the present calculation, $c = 10$, $r = .25$, and $s = 10$.

¹⁷ The general validity of the claim that a given percentage increase in the probability raises deterrence more than the same percentage increase in sentence length follows from note 15 above and the fact that, when a constant stream of disutility is discounted, $d(s)$ rises less than in proportion to sentence length (see note 6 above and accompanying text).

¹⁸ There has been considerable informal discussion in the criminological literature of the question addressed in this section. See, for example, Wilson and Herrnstein, at 397-401 (cited in note 5), and the sources cited therein. This literature emphasizes the view that raising the probability is more effective than raising the sanction. Support for this proposition also can be found in economically-oriented studies of crime. See, for example, Jeffrey Grogger, *Certainty Vs. Severity of Punishment*, 29 *Economic Inquiry* 297 (1991). Additionally, Gary Becker briefly examines the relative effectiveness of the probability versus the magnitude of sanctions and notes that the result depends on an individual's attitude toward risk. See Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 *J Political Economy* 169, 178 (1968). Michael Block and Robert Lind assume that individuals are risk preferrers in imprisonment and show that this implies that deterrence is enhanced more by raising the probability than raising the sanction. See Michael K. Block and Robert C. Lind, *An Economic Analysis of Crimes Punishable by Imprisonment*, 4 *J Legal Studies* 479, 481, 489-90 (1975).

4. Imprisonment Sanctions and Optimal Deterrence

In this section we examine the socially optimal magnitude and probability of imprisonment sanctions. We assume that the state can vary the imprisonment sanction up to some maximum sentence, and can alter the probability of apprehension. Obviously, the greater the use of prisons and the higher the probability of apprehension, the greater the costs incurred by society.

In choosing the sentence length and the probability of apprehension, we assume that society's objective is to minimize total social costs: the net harm generated by individuals' acts¹⁹ plus the various costs associated with the imposition of imprisonment sanctions -- the enforcement costs of apprehending individuals with probability p , the public costs of operating prisons, and the disutility suffered by those who are imprisoned.²⁰

For expositional simplicity, we initially discuss how the three forms of disutility from imprisonment affect the optimal system of deterrence in the absence of discounting, and then we consider the implications of discounting.

Our first claim is that *optimal imprisonment sanctions are maximal if individuals' total disutility rises proportionally with the length of the sentence -- that is, if individuals are risk neutral in imprisonment.*²¹ To explain, we will show that if the sanction is not maximal, total social costs can be lowered by increasing the sanction and reducing the probability of apprehension in such a way that deterrence is unchanged.

For example, suppose that the sentence is 4 years, that the maximum possible sentence is higher, say 30 years, that the disutility per year in jail is 3, and that the probability of apprehension is 30%. Consider increasing the sentence to 5 years and simultaneously lowering the probability so that the expected disutility of imprisonment, and thus deterrence, is constant. Since expected disutility initially is 3.6 ($= .3 \times 4 \times 3$), the probability must fall to 24% if expected disutility is to remain constant (since $.24 \times 5 \times 3 = 3.6$). Because this change leaves deterrence unaltered, the same number of individuals choose to commit harmful acts, so that net harm is

¹⁹ The net harm is the harm that individuals cause less any benefits they obtain that count in the social calculus (for example, a rapist's gains might not be given any weight).

²⁰ Our conclusions do not depend on the inclusion of the last category of social cost, the disutility of those in prison.

unchanged. Also, two other components of social cost are unchanged: the disutility suffered by punished individuals is not affected because the same number of individuals commit harmful acts and the expected disutility of imprisonment is the same as before; and the public costs of operating prisons is unchanged because the same number of individuals commit harmful acts and the expected prison term has not changed (it remains at 1.2 years since $.3 \times 4 \text{ years} = .24 \times 5 \text{ years} = 1.2 \text{ years}$). But because the probability of apprehension is lower, 24% instead of 30%, enforcement costs fall. Since enforcement costs fall and all other social costs are not affected by the increase in the sanction s and simultaneous reduction in the probability p that leaves deterrence constant, the change lowers total social costs and is socially beneficial.

It is clear that the preceding argument can be made whenever s is less than maximal, so that it must be socially desirable for s to be 30 years, the maximum imprisonment sentence. When s is 30 years, the probability of imprisonment will have been lowered from its original contemplated level. The optimal probability depends on how enforcement costs vary with p , on the level of harm from the criminal acts, and on the distribution of benefits from committing the acts among the population of potential violators.²²

We next show that *optimal imprisonment sanctions also are maximal if individuals' total disutility rises more than in proportion to the length of the sentence -- if individuals are risk averse in imprisonment*. The demonstration in this case is similar to that above, but involves two new elements which make the advantage of maximal sanctions even greater than in the previous case.

Again, suppose initially that the sentence is 4 years and that the probability of apprehension is 30%, but now assume that the disutility of sanctions is equal to the sentence length squared. Increase the sentence to 5 years and simultaneously reduce the probability so that the expected disutility of sanctions, and thus deterrence, is constant. The expected disutility originally was 4.8 ($= .3 \times 4^2$). For the expected disutility to be the same when the sentence is

²¹ See the Appendix for the general proof of this and the other claims made in this section.

²² There is not much that can be said in general about the character of the optimal probability of apprehension. It could be optimal for p to be relatively low, for example, if the cost of apprehension is high. Or it could be optimal for p to be relatively high, for instance, if the public cost of operating prisons is high (for if it is possible to deter most criminal acts by setting p high, there would be very few criminals who would be put in jail). For a formal description of the determination of the optimal probability of apprehension, see note 34 below.

raised to 5 years, the new probability of enforcement must decline to 19.2% (since $.192 \times 5^2 = 4.8$). Observe that the probability now falls *more* than in the risk-neutral case, where it declined to 24%. The reason is that here, in the risk-averse case, raising the sentence to 5 years creates more than a proportional increase in disutility;²³ hence, the probability can be lowered more than proportionally and still maintain deterrence.

Because the probability is lowered to 19.2% rather than to 24%, *the enforcement cost saving is greater* than it was in the risk-neutral case. Moreover, *there now is a saving in the operating costs of prisons*. This is for a similar reason. Specifically, because the probability falls more than in the risk-neutral case, the expected number of years criminals spend in prison declines even though deterrence is unchanged: the expected number of years per violator originally was 1.2 years ($= .3 \times 4$ years); now it is 0.96 years ($= .192 \times 5$ years). Since the number of violators is unchanged, the total number of person-years in prison falls, meaning that operating costs of prisons fall.

The remaining components of social costs -- the net harm and the expected disutility suffered by violators -- do not change (for the same reason as in the risk-neutral case). In sum, therefore, there are savings achieved by raising the sentence and lowering the probability, meaning that it must be optimal again for the sentence to be maximal.

Let us now indicate why *optimal imprisonment sanctions might not be maximal if individuals' total disutility rises less than proportionally with the imprisonment sentence -- if individuals are risk preferring in imprisonment*.

Suppose once more that the sentence is 4 years and that the probability is 30%, but now assume that the disutility from the first year in prison is 200 and that disutility per year thereafter is 10. If the sentence is raised to 5 years and the probability reduced so that expected disutility, and thus deterrence, is constant, the probability must become 28.75%. (This follows because $.3 \times [200 + (3 \times 10)] = .2875 \times [200 + (4 \times 10)]$.) Note that the probability cannot be lowered by as much as in the risk-neutral case, where it declined to 24%. Thus, the *enforcement cost saving, while positive, is less than in the risk-neutral case*. Moreover, unlike in the risk-neutral case, the

²³ The sentence rises from 4 years to 5 years, a 25% increase, but the disutility increases from 16 to 25, a 56% change.

expected imprisonment term now *rises*: it increases from 1.2 years ($= .3 \times 4$ years) to 1.44 years ($= .2875 \times 5$ years). Consequently, *the operating costs of prisons increase*. If these additional costs exceed the savings in enforcement costs, increasing the sentence and reducing the probability so as to maintain deterrence will *increase* total social costs.

The preceding observations suggest that, in the risk-preferring case, a less-than-maximal imprisonment sentence accompanied by a substantial probability may well be optimal.²⁴ The reason, in essence, is that the disutility per unit of sanctions is greatest when sanctions are not high (because disutility rises less than in proportion to the sentence); this implies that, to achieve a given expected disutility, it may be less costly to impose a moderate sentence with a substantial probability even though doing so means spending more on enforcement than otherwise.

Finally, consider the situation when individuals discount the future disutility of imprisonment and the state discounts the future costs of operating prisons. In general, the two discount rates may differ. Perhaps the most plausible assumption about the relationship between the rates is that the individual rate of discount of disutility exceeds the social rate of discount because, as noted earlier,²⁵ it is thought that criminals tend to be excessively present-oriented. We consider this assumption first (together with the simplifying assumption that the disutility of imprisonment is a constant amount per year).

If individuals discount disutility at a higher rate than society discounts imprisonment costs, optimal imprisonment sanctions might not be maximal. Suppose that the sentence is 4 years, the probability is 30%, the disutility of imprisonment each year is 20, the individual discount rate is 40%, the annual cost of imprisonment per person is \$30,000, and the social discount rate is 5%. If the sanction is raised to 5 years and the probability reduced so that the expected discounted disutility of sanctions is constant, the probability must decline to 27.26%.²⁶ Thus, *there is a reduction in enforcement costs*. The expected imprisonment term rises, from 1.2 years ($= .3 \times 4$ years) to 1.36 years ($= .2726 \times 5$ years), but prison costs are discounted. It can be

²⁴ See the Appendix for an example in which the optimal sentence is less than maximal when individuals are risk preferring in imprisonment.

²⁵ See note 5 above and the accompanying text.

²⁶ When $s = 4$, $c = 20$, and $r = .4$, the formula in note 16 above can be used to calculate that the present value of the disutility from the sentence is 51.78. Given $p = .3$, the expected discounted disutility therefore is 15.53 ($= .3 \times 51.78$). If $s = 5$, the present value of the disutility from the sentence becomes 56.98. For the expected

calculated that *the expected discounted costs of operating prisons rise*, from \$33,509 to \$37,177.²⁷ If this increase in the costs of prisons exceeds the savings in enforcement costs, total social costs do not fall, and again the optimal sanction may be less than maximal.

However, *if individuals discount disutility at a lower rate than society discounts imprisonment costs (or if the two rates are the same), optimal imprisonment sanctions are maximal*. This is because, when the sentence is raised and the probability lowered so as to keep deterrence constant, even though the expected prison sentence increases, the *discounted costs* of operating prisons fall (or remain the same if the discount rates are equal); thus, total social costs decline. In the example in the preceding paragraph, suppose that the discount rates are reversed -- now let the private rate be 5% and the social rate be 40%. Then the expected discounted costs of operating prisons are \$23,300 originally and fall to \$21,002 after the sentence is raised and the probability is lowered.²⁸

In general, when there is discounting and the disutility from imprisonment is not constant, we can obtain a qualitative understanding of optimal policy from the cases considered above. For example, suppose that disutility is high during the first year of imprisonment and then is roughly constant or falling, and that the individual rate of discount exceeds the social rate. Then the risk-preference effect and the discounting effect reinforce each other, making less-than-maximal sanctions more desirable than in either case alone. If, however, the disutility from imprisonment rises over time and the individual and social rates of discount are approximately equal, then maximal sanctions are likely to be optimal.²⁹

disutility to be the same, p must equal .2726 (since $.2726 \times 56.98 = 15.53$).

²⁷ Applying the formula in note 16 above with $s = 4$, $c = \$30,000$ (now interpreting c as the annual cost of operating prisons rather than the annual level of disutility), and $r = .05$, gives \$111,697. Since $p = .3$, expected discounted costs are \$33,509 ($= .3 \times \$111,697$). If $s = 5$ and $p = .2726$, the corresponding numbers are \$136,379 and \$37,177.

²⁸ The calculations are analogous to those performed previously, and are not reported here.

²⁹ The only result in this section that has been demonstrated previously is that the optimal imprisonment sanction is maximal when disutility is proportional to the sanction. See Steven Shavell, *Specific versus General Enforcement of Law*, 99 *J Political Economy* 1088, 1098 (1991). This conclusion also was implicit in Becker, at 183 (cited in note 18). Shavell stated, but did not prove, the results regarding optimal imprisonment sanctions when disutility rises more than or less than proportionally with the sanction. However, the preceding conclusions, including when disutility is proportional to the imprisonment sanction, are not generally appreciated. Moreover, Shavell and Becker ignored the issue of discounting (both private and social).

5. Concluding Comments

(a) *Relevance of results.* The results of this article will be of interest in settings in which there is information about the form of individuals' disutility from imprisonment or about the relationship between the private and social rates of discount. For instance, for individuals who commit white collar crimes, the disutility of being in prison at all may be substantial and the stigma and loss of earning power may depend relatively little on the length of imprisonment. Thus, such individuals are likely to be risk preferring in imprisonment, suggesting that less-than-maximal sanctions, combined with relatively high probabilities of apprehension, may be optimal. Similarly, it is plausible that young males who commit crimes discount the future disutility of imprisonment at a higher rate than the social discount rate, also suggesting that limited prison sentences and relatively high probabilities are optimal. In other contexts, the appropriate assumptions to make about the form of disutility from imprisonment and about discount rates may be different; as suggested by our remarks in Section 2, each of the assumptions we discussed probably has non-trivial empirical importance.

(b) *Relationship to analysis of monetary sanctions.* It is interesting to compare the results obtained here to those when sanctions are monetary. If sanctions are monetary and individuals are risk neutral in wealth, then, as was suggested by Gary Becker,³⁰ optimal sanctions are maximal; the argument used here to demonstrate the analogous result when the sanction is imprisonment and individuals are risk neutral with respect to sentence length is in essence Becker's argument.

If sanctions are monetary and individuals are risk averse, then, as shown by us previously,³¹ optimal sanctions may not be maximal, whereas here they are. A way to understand the different conclusions is to consider raising the sanction from a less-than-maximal level and lowering the probability so as to preserve deterrence. If sanctions are monetary, this change reduces fine revenue because expected sanctions fall, and the reduction in fine revenue lowers social welfare.³² But if the sanction is imprisonment, the change reduces the expected length of

³⁰ See Becker (cited in note 18).

³¹ See A. Mitchell Polinsky and Steven Shavell. *The Optimal Tradeoff Between the Probability and Magnitude of Fines*, 69 Amer Econ Rev 880 (1979).

³² The reduction in fine revenue reflects the increased disutility from the bearing of greater financial risk

prison sentences, rather than expected fines, which saves operating costs and thereby raises social welfare.

If sanctions are monetary and individuals are risk preferrers, then it can be shown that optimal sanctions must be maximal,³³ whereas here they may not be maximal. The reasons for this difference are analogous to those just stated in the risk-averse case.

when the sanction rises and the probability falls.

³³ The proof of this point may be sketched as follows. Let p^* be the optimal probability and s^* the optimal monetary sanction, and suppose initially that $s^* < s_m$, where s_m is the maximum possible sanction. Raise the sanction to s' and lower the probability to p' so that the expected disutility of sanctions is the same -- that is, so that $(1 - p^*)U(y) + p^*U(y - s^*) = (1 - p')U(y) + p'U(y - s')$, where y is the individual's initial income. Since $U'' > 0$ (risk preference), we know that $p's' > p^*s^*$, that is, fine revenue per violator increases. Because the number of violators has not changed, total fine revenue must increase. Further, since $p' < p^*$, enforcement costs fall. Therefore, government net revenue rises. Since the change in the sanction and the probability leaves deterrence and harm unaltered and does not affect the expected utility of those who violate the law, yet produces an increase in government net revenue, it must result in an increase in social welfare. Thus, the original $s^* < s_m$ could not have been optimal.

Appendix

In this appendix we formally demonstrate the results in Section 4 concerning the optimal magnitude of imprisonment sanctions.

We assume that the benefit that an individual obtains from committing an act that causes harm varies among individuals, that injurers are apprehended with some probability, that the imprisonment sentence imposed on them is constrained by some maximum, and that the state incurs enforcement costs to maintain the probability and to operate prisons. The following notation will be used:

- b = benefit that an individual obtains from committing the harmful act;
 $0 \leq b \leq B$;
- $f(b)$ = density of individuals whose benefit is b ;
- $F(b)$ = cumulative distribution of $f(b)$;
- h = harm from committing the act;
- s = length of prison sentence;
- s_m = maximum prison sentence;
- $d(s)$ = total disutility from sentence of length s (possibly discounted); $d(s) > 0$;
- c = constant disutility per year of imprisonment (risk-neutral case only);
- $t(s)$ = total public operating cost of prisons given sentence of length s (possibly discounted); $t'(s) > 0$;
- q = public operating cost of prison per person per year of imprisonment;
- p = probability of apprehension;
- $g(p)$ = public cost of maintaining the probability of apprehension at p ;
 $g'(p) > 0$.

An individual will commit the harmful act if his benefit equals or exceeds the expected disutility of the sanction:

$$b \geq pd(s). \quad (1)$$

The social objective is to minimize total social costs:

$$\int_{pd(s)}^B (h - b)f(b)db + \int_{pd(s)}^B pd(s)f(b)db + \int_{pd(s)}^B pt(s)f(b)db + g(p). \quad (2)$$

The first term is the net harm caused; the second is the expected disutility suffered by those punished; the third term is the cost of operating prisons; the last is the enforcement cost.

The social problem is to choose p and s to minimize (2); the optimal values will be denoted by p^* and s^* . We assume that $p^* > 0$ (otherwise the problem of enforcement is of no interest).

We now discuss optimal deterrence in the five cases considered in the text, the first three of which focus on the form of disutility from imprisonment and the last two of which consider discounting. Because our principal interest is in s^* , we will comment on the determination of p^* only in a footnote (there is not much of interest that can be said about its characterization).

Optimal Imprisonment Sanctions when Individuals' Total Disutility Rises Proportionally with the Length of the Sentence -- the Risk-Neutral Case. In this case, $d(s) = cs$ and, because discounting is being ignored for now, $t(s) = qs$. Assume that $s^* < s_m$, the maximal sanction. Raise s to a higher s' (which is possible since $s^* < s_m$). Choose p' such that $p's' = p^*s^*$. Then $p^*d(s^*) = cp^*s^* = cp's' = p'd(s')$. Similarly, $p^*t(s^*) = p't(s')$. Hence, from inspection of (2), it is apparent that the first three terms do not change, and that the last term falls. Thus, $s^* < s_m$ could not have been optimal; $s^* = s_m$ must hold.³⁴

Optimal Imprisonment Sanctions when Individuals' Total Disutility Rises More than in Proportion to the Length of the Sentence -- the Risk-Averse Case. In this case $d''(s) > 0$ and $t(s) = qs$. Assume that $s^* < s_m$. Raise s to s' and choose a lower p' such that $p'd(s') = p^*d(s^*)$. From inspection of (2), it is apparent that the first two terms do not change. Moreover, because $d''(s) > 0$, we know that $p's' < p^*s^*$, which implies that the third term declines. Clearly, the fourth term also falls. Hence, it must be that $s^* = s_m$.

³⁴ To determine the optimal probability in the present case, rewrite (2) using the fact that $d(s) = cs$, $t(s) = qs$, and $s^* = s_m$. Differentiating the resulting expression with respect to p and setting the result equal to zero gives the first-order condition, which can be rewritten as

$$cs_m(h + qs_m)f(pcs_m) - (c + q)s_m[1 - F(pcs_m)] = g'(p). \quad (i)$$

The left-hand side of (i) is the marginal benefit of raising p , which reflects the reduction in harm and in the public cost of imprisonment due to greater deterrence, less the greater disutility and imprisonment costs due to the apprehension of a higher fraction of those who commit the harmful act. The right-hand side of (i) is the marginal cost of raising p . As observed in note 22 above, the optimal probability could be relatively low or relatively high. (Because the first-order conditions in the other cases are similar, we will not comment further on the determination of the optimal probability.)

Optimal Imprisonment Sanctions when Individuals' Total Disutility Rises Less than Proportionally with the Length of the Sentence -- the Risk-Preferring Case. In this case, $d''(s) < 0$ and $t(s) = qs$. As observed in text, the type of argument used in the two previous cases does not apply. Specifically, suppose that $s^* < s_m$ and that we raise s to s' and choose a lower p' such that $p'd(s') = p^*d(s^*)$. Again, the first two terms of (2) do not change and the last term declines. But $d''(s) < 0$ implies that $p's' > p^*s^*$. Thus, the third term rises, so that we cannot conclude that total social costs fall. To see that the optimal prison sentence may be less than maximal in the present case, consider the following example. Let $d(s) = 200 + 10(s - 1)$, $s_m = 30$ years, $g(p) = \$20p$, $h = \$50$, $q = \$20$, and the distribution of benefits be $f(i) = .02$ for $i = \$1, \dots, \20 , $f(i) = .01$ for $i = \$21, \dots, \30 , $f(i) = 0$ for $i = \$31, \dots, \70 , $f(i) = .01$ for $i = \$71, \dots, \80 , and $f(i) = .02$ for $i = \$81, \dots, \100 . In this example, the optimal prison sentence is one year, far below the maximum, and the optimal probability of apprehension is 11%.³⁵

In the two remaining cases, we consider discounting and assume that the disutility of imprisonment is a constant amount, c , per year. Let:

a = rate at which individuals discount disutility;

r = rate at which society discounts the operating costs of prisons.

Optimal Imprisonment Sanctions when Individuals Discount Disutility at a Higher Rate than Society Discounts Imprisonment Costs. This case parallels the case of risk preference just considered. Here, $d(s) = c[1 + 1/(1+a) + \dots + 1/(1+a)^{s-1}]$ and $t(s) = q[1 + 1/(1+r) + 1/(1+r)^2 + \dots + 1/(1+r)^{s-1}]$. Suppose that $s^* < s_m$ and that we raise s to s' and choose a lower p' such that $p'd(s') = p^*d(s^*)$. As before, the first two terms of (2) are unaffected and the last term falls. To prove that $p't(s') > p^*t(s^*)$, and therefore that the third term rises, it suffices to show that $d(s^*)/d(s') > t(s^*)/t(s')$ (for from $p'd(s') = p^*d(s^*)$, we have $p' = p^*d(s^*)/d(s')$). But the latter follows from the assumption in the present case that $a > r$.³⁶ Thus, the third term rises, so that

³⁵ The optimal values were determined by numerical calculation, allowing s to take on only discrete values and p to be a discrete percentage between 0 and 100.

³⁶ Suppose, say, $s' = s^* + 1$. Then $d(s^*)/d(s') = [1 + 1/(1+a) + \dots + 1/(1+a)^{s^*-1}] / [1 + 1/(1+a) + \dots + 1/(1+a)^{s^*}]$ and $t(s^*)/t(s') = [1 + 1/(1+r) + \dots + 1/(1+r)^{s^*-1}] / [1 + 1/(1+r) + \dots + 1/(1+r)^{s^*}]$. To demonstrate that the former exceeds the latter, it is clearly sufficient to show in general that the ratio $[1 + z + z^2 + \dots + z^{s-1}] / [1 + z + z^2 + \dots + z^s]$ is decreasing in z . To prove this, it in turn suffices to demonstrate that the numerator of the derivative with respect to z of the ratio is negative (since the denominator of the derivative is positive). The

we cannot conclude that total social costs fall. An example in which the optimal sentence is less than maximal is as follows. Let $c = 100$, $a = .2$, $q = \$10$, $r = .05$, $s_m = 30$ years, $g(p) = \$20p$, $h = \$50$, and the distribution of benefits be as in the example in the risk-preferring case above. Then the optimal sentence is 10 years, and the optimal probability is 4%.³⁷

Optimal Imprisonment Sanctions when Individuals Discount Disutility at a Lower Rate than Society Discounts Imprisonment Costs (or at the Same Rate). This case parallels the case of risk aversion. Assume that $s^* < s_m$, raise s to s' , and choose a lower p' such that $p'd(s') = p*d(s^*)$. The first two terms of (2) do not change and the fourth term declines. To see that $p't(s') < p*t(s^*)$, so that the third term also falls, it suffices to show that $d(s^*)/d(s') < t(s^*)/t(s')$, which follows from the present assumption that $a < r$.³⁸ Hence, $s^* = s_m$ must hold. (If $a = r$, the third term is not affected, but the argument works because the fourth term still declines.)

numerator of the derivative is $[1 + 2z + 3z^2 + \dots + (s - 1)z^{s-2}][1 + z + z^2 + \dots + z^s] - [1 + z + z^2 + \dots + z^{s-1}][1 + 2z + 3z^2 + \dots + sz^{s-1}]$, which after cancellation and expansion equals $[z^s + 2z^{s+1} + \dots + (s - 1)z^{2s-2}] - [sz^s + sz^{s+1} + \dots + sz^{2s-2}] - sz^{s-1}$. The latter is negative, since the first term in brackets is less than the second term in brackets.

³⁷ This was demonstrated by numerical calculation in the way described in note 35 above.

³⁸ This result flows immediately from what was demonstrated in note 36 above (now given $a < r$).