Crime and Punishment: An Economic Approach

Gary S. Becker*

Columbia University

I. Introduction

Since the turn of the century, legislation in Western countries has expanded rapidly to reverse the brief dominance of laissez faire during the nineteenth century. The state no longer merely protects against violations of person and property through murder, rape, or burglary but also restricts "discrimination" against certain minorities, collusive business arrangements, "jaywalking," travel, the materials used in construction, and thousands of other activities. The activities restricted not only are numerous but also range widely, affecting persons in very different pursuits and of diverse social backgrounds, education levels, ages, races, etc. Moreover, the likelihood that an offender will be discovered and convicted and the nature and extent of punishments differ greatly from person to person and activity to activity. Yet, in spite of such diversity, some common properties are shared by practically all legislation, and these properties form the subject matter of this essay.

In the first place, obedience to law is not taken for granted, and public and private resources are generally spent in order both to prevent offenses and to apprehend offenders. In the second place, conviction is not generally considered sufficient punishment in itself; additional and sometimes severe punishments are meted out to those convicted. What determines the amount and type of resources and punishments used to enforce a piece of legislation? In particular, why does enforcement differ so greatly among different kinds of legislation?

* I would like to thank the Lilly Endowment for financing a very productive summer in 1965 at the University of California at Los Angeles. While there I received very helpful comments on an earlier draft from, among others, Armen Alchian, Roland McKean, Harold Demsetz, Jack Hirshliefer, William Meckling, Gordon Tullock, and Oliver Williamson. I have also benefited from comments received at seminars at the University of Chicago, Hebrew University, RAND Corporation, and several times at the Labor Workshop of Columbia; assistance and suggestions from Isaac Ehrlich and Robert Michael; and suggestions from the editor of this journal.

Copyright © 2001. All Rights Reseved.

The main purpose of this essay is to answer normative versions of these questions, namely, how many resources and how much punishment *should* be used to enforce different kinds of legislation? Put equivalently, although more strangely, how many offenses *should* be permitted and how many offenders *should* go unpunished? The method used formulates a measure of the social loss from offenses and finds those expenditures of resources and punishments that minimize this loss. The general criterion of social loss is shown to incorporate as special cases, valid under special assumptions, the criteria of vengeance, deterrence, compensation, and rehabilitation that historically have figured so prominently in practice and criminological literature.

The optimal amount of enforcement is shown to depend on, among other things, the cost of catching and convicting offenders, the nature of punishments—for example, whether they are fines or prison terms—and the responses of offenders to changes in enforcement. The discussion, therefore, inevitably enters into issues in penology and theories of criminal behavior. A second, although because of lack of space subsidiary, aim of this essay is to see what insights into these questions are provided by our "economic" approach. It is suggested, for example, that a useful theory of criminal behavior can dispense with special theories of anomie, psychological inadequacies, or inheritance of special traits and simply extend the economist's usual analysis of choice.

II. Basic Analysis

A. The Cost of Crime

Although the word "crime" is used in the title to minimize terminological innovations, the analysis is intended to be sufficiently general to cover all violations, not just felonies—like murder, robbery, and assault, which receive so much newspaper coverage—but also tax evasion, the so-called white-collar crimes, and traffic and other violations. Looked at this broadly, "crime" is an economically important activity or "industry," notwithstanding the almost total neglect by economists.[1] Some relevant evidence recently put together by the President's Commission on Law

[1] This neglect probably resulted from an attitude that illegal activity is too immoral to merit any systematic scientific attention. The influence of moral attitudes on a scientific analysis is seen most clearly in a discussion by Alfred Marshall. After arguing that even fair gambling is an "economic blunder" because of diminishing marginal utility, he says, "It is true that this loss of probable happiness need not be greater than the pleasure derived from the excitement of gambling, and we are then thrown back upon the induction [*sic*] that pleasures of gambling are in Bentham's phrase 'impure'; since experience shows that they are likely to engender a restless, feverish character, unsuited for steady work as well as for the higher and more solid pleasures of life" (Marshall, 1961, Note X, Mathematical Appendix).

Copyright © 2001. All Rights Reserved.

Enforcement and Administration of Justice (the "Crime Commission") is reproduced in Table 1. Public expenditures in 1965 at the federal, state, and local levels on police, criminal courts and counsel, and "corrections" amounted to over $4 billion, while private outlays on burglar alarms, guards, counsel, and some other forms of protection were about $2 billion. Unquestionably, public and especially private expenditures are significantly understated, since expenditures by many public agencies in the course of enforcing particular pieces of legislation, such as state fair-employment laws,[2] are not included, and a myriad of private precautions against crime, ranging from suburban living to taxis, are also excluded.

TABLE 1
ECONOMIC COSTS OF CRIMES

Type	Costs (Millions of Dollars)
Crimes against persons	815
Crimes against property	3,932
Illegal goods and services	8,075
Some other crimes	2,036
Total	14,858
Public expenditures on police, prosecution, and courts	3,178
Corrections	1,034
Some private costs of combatting crime	1,910
Over-all total	20,980

Source: President's Commission, (1967*d*, p. 44).

Table 1 also lists the Crime Commission's estimates of the direct costs of various crimes. The gross income from expenditures on various kinds of illegal consumption, including narcotics, prostitution, and mainly gambling, amounted to over $8 billion. The value of crimes against property, including fraud, vandalism, and theft, amounted to almost $4 billion,[3] while about $3 billion worth resulted from the loss of earnings due to homicide, assault, or other crimes. All the costs listed in the table total about $21 billion, which is almost 4 per cent of reported national

[2] Expenditures by the thirteen states with such legislation in 1959 totaled almost $2 million (see Landes, 1966).

[3] Superficially, frauds, thefts, etc., do not involve true social costs but are simply transfers, with the loss to victims being compensated by equal gains to criminals. While these are transfers, their market value is, nevertheless, a first approximation to the direct social cost. If the theft or fraud industry is "competitive," the sum of the value of the criminals' time input—including the time of "fences" and prospective time in prison—plus the value of capital input, compensation for risk, etc., would approximately equal the market value of the loss to victims. Consequently, aside from the input of intermediate products, losses can be taken as a measure of the value of the labor and capital input into these crimes, which are true social costs.

Copyright © 2001. All Rights Reseved.

income in 1965. If the sizeable omissions were included, the percentage might be considerably higher.

Crime has probably become more important during the last forty years. The Crime Commission presents no evidence on trends in costs but does present evidence suggesting that the number of major felonies per capita has grown since the early thirties (President's Commission, 1967*a*, pp. 22–31). Moreover, with the large growth of tax and other legislation, tax evasion and other kinds of white-collar crime have presumably grown much more rapidly than felonies. One piece of indirect evidence on the growth of crime is the large increase in the amount of currency in circulation since 1929. For sixty years prior to that date, the ratio of currency either to all money or to consumer expenditures had declined very substantially. Since then, in spite of further urbanization and income growth and the spread of credit cards and other kinds of credit,[4] both ratios have increased sizeably.[5] This reversal can be explained by an unusual increase in illegal activity, since currency has obvious advantages over checks in illegal transactions (the opposite is true for legal transactions) because no record of a transaction remains.[6]

B. The Model

It is useful in determining how to combat crime in an optimal fashion to develop a model to incorporate the behavioral relations behind the costs listed in Table 1. These can be divided into five categories: the relations between (1) the number of crimes, called "offenses" in this essay, and the cost of offenses, (2) the number of offenses and the punishments meted out, (3) the number of offenses, arrests, and convictions and the public expenditures on police and courts, (4) the number of convictions and the costs of imprisonments or other kinds of punishments, and (5) the number of offenses and the private expenditures on protection and apprehension. The first four are discussed in turn, while the fifth is postponed until a later section.

1. Damages

Usually a belief that other members of society are harmed is the motivation behind outlawing or otherwise restricting an activity. The amount of harm

[4] For an analysis of the secular decline to 1929 that stresses urbanization and the growth in incomes, see Cagan (1965, chap. iv).

[5] In 1965, the ratio of currency outstanding to consumer expenditures was 0.08, compared to only 0.05 in 1929. In 1965, currency outstanding per family was a whopping $738.

[6] Cagan (1965, chap. iv) attributes much of the increase in currency holdings between 1929 and 1960 to increased tax evasion resulting from the increase in tax rates.

Copyright © 2001. All Rights Reserved.

would tend to increase with the activity level, as in the relation

$$H_i = H_i(O_i),$$

with (1)

$$H_i' = \frac{dH_i}{dO_i} > 0,$$

where H_i is the harm from the ith activity and O_i is the activity level.[7] The concept of harm and the function relating its amount to the activity level are familiar to economists from their many discussions of activities causing external diseconomies. From this perspective, criminal activities are an important subset of the class of activities that cause diseconomies, with the level of criminal activities measured by the number of offenses.

The social value of the gain to offenders presumably also tends to increase with the number of offenses, as in

$$G = G(O),$$

with (2)

$$G' = \frac{dG}{dO} > 0.$$

The net cost or damage to society is simply the difference between the harm and gain and can be written as

$$D(O) = H(O) - G(O). \tag{3}$$

If, as seems plausible, offenders usually eventually receive diminishing marginal gains and cause increasing marginal harm from additional offenses, $G'' < 0$, $H'' > 0$, and

$$D'' = H'' - G'' > 0, \tag{4}$$

which is an important condition used later in the analysis of optimality positions (see, for example, the Mathematical Appendix). Since both H' and $G' > 0$, the sign of D' depends on their relative magnitudes. It follows from (4), however, that

$$D'(O) > 0 \text{ for all } O > O_a \text{ if } D'(O_a) \geq 0. \tag{5}$$

Until Section V the discussion is restricted to the region where $D' > 0$, the region providing the strongest justification for outlawing an activity. In that section the general problem of external diseconomies is reconsidered from our viewpoint, and there $D' < 0$ is also permitted.

The top part of Table 1 lists costs of various crimes, which have been interpreted by us as estimates of the value of resources used up in these

[7] The ith subscript will be suppressed whenever it is to be understood that only one activity is being discussed.

Copyright © 2001. All Rights Reseved.

crimes. These values are important components of, but are not identical to, the net damages to society. For example, the cost of murder is measured by the loss in earnings of victims and excludes, among other things, the value placed by society on life itself; the cost of gambling excludes both the utility to those gambling and the "external" disutility to some clergy and others; the cost of "transfers" like burglary and embezzlement excludes social attitudes toward forced wealth redistributions and also the effects on capital accumulation of the possibility of theft. Consequently, the $15 billion estimate for the cost of crime in Table 1 may be a significant understatement of the net damages to society, not only because the costs of many white-collar crimes are omitted, but also because much of the damage is omitted even for the crimes covered.

2. The Cost of Apprehension and Conviction

The more that is spent on policemen, court personnel, and specialized equipment, the easier it is to discover offenses and convict offenders. One can postulate a relation between the output of police and court "activity" and various inputs of manpower, materials, and capital, as in $A = f(m, r, c)$, where f is a production function summarizing the "state of the arts." Given f and input prices, increased "activity" would be more costly, as summarized by the relation

$$C = C(A)$$

and (6)

$$C' = \frac{dC}{dA} > 0.$$

It would be cheaper to achieve any given level of activity the cheaper were policemen,[8] judges, counsel, and juries and the more highly developed the state of the arts, as determined by technologies like fingerprinting, wiretapping, computer control, and lie-detecting.[9]

One approximation to an empirical measure of "activity" is the number of offenses cleared by conviction. It can be written as

$$A \cong pO, \tag{7}$$

where p, the ratio of offenses cleared by convictions to all offenses, is the over-all probability that an offense is cleared by conviction. By substituting

[8] According to the Crime Commission, 85–90 per cent of all police costs consist of wages and salaries (President's Commission, 1967*a*, p. 35).

[9] A task-force report by the Crime Commission deals with suggestions for greater and more efficient usage of advanced technologies (President's Commission, 1967*e*).

Copyright © 2001. All Rights Reserved.

(7) into (6) and differentiating, one has

$$C_p = \frac{\partial C(pO)}{\partial p} = C'O > 0$$

and (8)

$$C_o = C'p > 0$$

if $pO \neq 0$. An increase in either the probability of conviction or the number of offenses would increase total costs. If the marginal cost of increased "activity" were rising, further implications would be that

$$\begin{aligned} C_{pp} &= C''O^2 > 0, \\ C_{oo} &= C''p^2 > 0, \end{aligned} \tag{9}$$

and

$$C_{po} = C_{op} = C''pO + C' > 0.$$

A more sophisticated and realistic approach drops the implication of (7) that convictions alone measure "activity," or even that p and O have identical elasticities, and introduces the more general relation

$$A = h(p, O, a). \tag{10}$$

The variable a stands for arrests and other determinants of "activity," and there is no presumption that the elasticity of h with respect to p equals that with respect to O. Substitution yields the cost function $C = C(p, O, a)$. If, as is extremely likely, h_p, h_o, and h_a are all greater than zero, then clearly C_p, C_o, and C_a are all greater than zero.

In order to insure that optimality positions do not lie at "corners," it is necessary to place some restrictions on the second derivatives of the cost function. Combined with some other assumptions, it is *sufficient* that

$$\begin{aligned} C_{pp} &\geq 0, \\ C_{oo} &\geq 0, \end{aligned} \tag{11}$$

and

$$C_{po} \cong 0$$

(see the Mathematical Appendix). The first two restrictions are rather plausible, the third much less so.[10]

Table 1 indicates that in 1965 public expenditures in the United States on police and courts totaled more than $3 billion, by no means a minor

[10] Differentiating the cost function yields $C_{pp} = C''(h_p)^2 + C'h_{pp}$; $C_{oo} = C''(h_o)^2 + C'h_{oo}$; $C_{po} = C''h_oh_p + C'h_{po}$. If marginal costs were rising, C_{pp} or C_{oo} could be negative only if h_{pp} or h_{oo} were sufficiently negative, which is not very likely. However, C_{po} would be approximately zero only if h_{po} were sufficiently negative, which is also unlikely. Note that if "activity" is measured by convictions alone, $h_{pp} = h_{oo} = 0$, and $h_{po} > 0$.

Copyright © 2001. All Rights Reseved.

item. Separate estimates were prepared for each of seven major felonies.[11] Expenditures on them averaged about $500 per offense (reported) and about $2,000 per person arrested, with almost $1,000 being spent per murder (President's Commission, 1967*a*, pp. 264–65); $500 is an estimate of the average cost

$$AC = \frac{C(p, O, a)}{O}$$

of these felonies and would presumably be a larger figure if the number of either arrests or convictions were greater. Marginal costs (C_o) would be at least $500 if condition (11), $C_{oo} \geq 0$, were assumed to hold throughout.

3. The Supply of Offenses

Theories about the determinants of the number of offenses differ greatly, from emphasis on skull types and biological inheritance to family upbringing and disenchantment with society. Practically all the diverse theories agree, however, that when other variables are held constant, an increase in a person's probability of conviction or punishment if convicted would generally decrease, perhaps substantially, perhaps negligibly, the number of offenses he commits. In addition, a common generalization by persons with judicial experience is that a change in the probability has a greater effect on the number of offenses than a change in the punishment,[12] although, as far as I can tell, none of the prominent theories shed any light on this relation.

The approach taken here follows the economists' usual analysis of choice and assumes that a person commits an offense if the expected utility to him exceeds the utility he could get by using his time and other resources at other activities. Some persons become "criminals," therefore, not because their basic motivation differs from that of other persons, but because their benefits and costs differ. I cannot pause to discuss the many general implications of this approach,[13] except to remark that criminal behavior becomes part of a much more general theory and does not require ad hoc concepts of differential association, anomie, and the like,[14] nor does it assume perfect knowledge, lightening-fast calculation, or any of the other caricatures of economic theory.

[11] They are willful homicide, forcible rape, robbery, aggravated assault, burglary, larceny, and auto theft.

[12] For example, Lord Shawness (1965) said, "Some judges preoccupy themselves with methods of punishment. This is their job. But in preventing crime it is of less significance than they like to think. Certainty of detection is far more important than severity of punishment." Also see the discussion of the ideas of C. B. Beccaria, an insightful eighteenth-century Italian economist and criminologist, in Radzinowicz (1948, I, 282).

[13] See, however, the discussions in Smigel (1965) and Ehrlich (1967).

[14] For a discussion of these concepts, see Sutherland (1960).

Copyright © 2001. All Rights Reserved.

This approach implies that there is a function relating the number of offenses by any person to his probability of conviction, to his punishment if convicted, and to other variables, such as the income available to him in legal and other illegal activities, the frequency of nuisance arrests, and his willingness to commit an illegal act. This can be represented as

$$O_j = O_j(p_j, f_j, u_j), \tag{12}$$

where O_j is the number of offenses he would commit during a particular period, p_j his probability of conviction per offense, f_j his punishment per offense, and u_j a portmanteau variable representing all these other influences.[15]

Since only convicted offenders are punished, in effect there is "price discrimination" and uncertainty: if convicted, he pays f_j per convicted offense, while otherwise he does not. An increase in either p_j or f_j would reduce the utility expected from an offense and thus would tend to reduce the number of offenses because either the probability of "paying" the higher "price" or the "price" itself would increase.[16] That is,

$$O_{p_j} = \frac{\partial O_j}{\partial p_j} < 0$$

and (13)

$$O_{f_j} = \frac{\partial O_j}{\partial f_j} < 0,$$

which are the generally accepted restrictions mentioned above. The effect of changes in some components of u_j could also be anticipated. For example, a rise in the income available in legal activities or an increase in law-abidingness due, say, to "education" would reduce the incentive to enter illegal activities and thus would reduce the number of offenses. Or a shift in the form of the punishment, say, from a fine to imprisonment,

[15] Both p_j and f_j might be considered distributions that depend on the judge, jury, prosecutor, etc., that j happens to receive. Among other things, u_j depends on the p's and f's meted out for other competing offenses. For evidence indicating that offenders do substitute among offenses, see Smigel (1965).

[16] The utility expected from committing an offense is defined as

$$EU_j = p_j U_j(Y_j - f_j) + (1 - p_j) U_j(Y_j),$$

where Y_j is his income, monetary plus psychic, from an offense; U_j is his utility function; and f_j is to be interpreted as the monetary equivalent of the punishment. Then

$$\frac{\partial EU_j}{\partial p_j} = U_j(Y_j - f_j) - U_j(Y_j) < 0$$

and

$$\frac{\partial EU_j}{\partial f_j} = -p_j U_j'(Y_j - f_j) < 0$$

as long as the marginal utility of income is positive. One could expand the analysis by incorporating the costs and probabilities of arrests, detentions, and trials that do not result in conviction.

Copyright © 2001. All Rights Reseved.

would tend to reduce the number of offenses, at least temporarily, because they cannot be committed while in prison.

This approach also has an interesting interpretation of the presumed greater response to a change in the probability than in the punishment. An increase in p_j "compensated" by an equal percentage reduction in f_j would not change the expected income from an offense[17] but could change the expected utility, because the amount of risk would change. It is easily shown that an increase in p_j would reduce the expected utility, and thus the number of offenses, more than an equal percentage increase in f_j[18] if j has preference for risk; the increase in f_j would have the greater effect if he has aversion to risk; and they would have the same effect if he is risk neutral.[19] The widespread generalization that offenders are more deterred by the probability of conviction than by the punishment when convicted turns out to imply in the expected-utility approach that offenders are risk preferrers, at least in the relevant region of punishments.

The total number of offenses is the sum of all the O_j and would depend on the set of p_j, f_j, and u_j. Although these variables are likely to differ significantly between persons because of differences in intelligence, age, education, previous offense history, wealth, family upbringing, etc., for simplicity I now consider only their average values, p, f, and u,[20] and write the market offense function as

$$O = O(p, f, u). \tag{14}$$

This function is assumed to have the same kinds of properties as the individual functions, in particular, to be negatively related to p and f and to be more responsive to the former than the latter if, and only if, offenders on balance have risk preference. Smigel (1965) and Ehrlich (1967) estimate

[17] $EY_j = p_j(Y_j - f_j) + (1 - p_j)Y_j = Y_j - p_j f_j$.

[18] This means that an increase in p_j "compensated" by a reduction in f_j would reduce utility and offenses.

[19] From n. 16

$$\frac{-\partial EU_j}{\partial p_j}\frac{p_j}{U_j} = [U_j(Y_j) - U_j(Y_j - f_j)]\frac{p_j}{U_j} \gtreqqless \frac{-\partial EU_j}{\partial f_j}\frac{f_j}{U_j} = p_j U_j'(Y_j - f_j)\frac{f_j}{U_j}$$

as

$$\frac{U_j(Y_j) - U_j(Y_j - f_j)}{f_j} \gtreqqless U_j'(Y_j - f_j).$$

The term on the left is the average change in utility between $Y_j - f_j$ and Y_j. It would be greater than, equal to, or less than $U_j'(Y_j - f_j)$ as $U_j'' \gtreqqless 0$. But risk preference is defined by $U_j'' > 0$, neutrality by $U_j'' = 0$, and aversion by $U_j'' < 0$.

[20] p can be defined as a weighted average of the p_j, as

$$p = \sum_{j=1}^{n} \frac{O_j p_j}{\sum_{i=1}^{n} O_i},$$

and similar definitions hold for f and u.

Copyright © 2001. All Rights Reserved.

functions like (14) for seven felonies reported by the Federal Bureau of Investigation using state data as the basic unit of observation. They find that the relations are quite stable, as evidenced by high correlation coefficients; that there are significant negative effects on O of p and f; and that usually the effect of p exceeds that of f, indicating preference for risk in the region of observation.

A well-known result states that, in equilibrium, the real incomes of persons in risky activities are, at the margin, relatively high or low as persons are generally risk avoiders or preferrers. If offenders were risk preferrers, this implies that the real income of offenders would be lower, at the margin, than the incomes they could receive in less risky legal activities, and conversely if they were risk avoiders. Whether "crime pays" is then an implication of the attitudes offenders have toward risk and is not directly related to the efficiency of the police or the amount spent on combatting crime. If, however, risk were preferred at some values of p and f and disliked at others, public policy could influence whether "crime pays" by its choice of p and f. Indeed, it is shown later that the social loss from illegal activities is usually minimized by selecting p and f in regions where risk is preferred, that is, in regions where "crime does not pay."

4. Punishments

Mankind has invented a variety of ingenious punishments to inflict on convicted offenders: death, torture, branding, fines, imprisonment, banishment, restrictions on movement and occupation, and loss of citizenship are just the more common ones. In the United States, less serious offenses are punished primarily by fines, supplemented occasionally by probation, petty restrictions like temporary suspension of one's driver's license, and imprisonment. The more serious offenses are punished by a combination of probation, imprisonment, parole, fines, and various restrictions on choice of occupation. A recent survey estimated for an average day in 1965 the number of persons who were either on probation, parole, or institutionalized in a jail or juvenile home (President's Commission 1967*b*). The total number of persons in one of these categories came to about 1,300,000, which is about 2 per cent of the labor force. About one-half were on probation, one-third were institutionalized, and the remaining one-sixth were on parole.

The cost of different punishments to an offender can be made comparable by converting them into their monetary equivalent or worth, which, of course, is directly measured only for fines. For example, the cost of an imprisonment is the discounted sum of the earnings foregone and the value placed on the restrictions in consumption and freedom. Since the earnings foregone and the value placed on prison restrictions vary from person to person, the cost even of a prison sentence of given duration is

Copyright © 2001. All Rights Reseved.

not a unique quantity but is generally greater, for example, to offenders who could earn more outside of prison.[21] The cost to each offender would be greater the longer the prison sentence, since both foregone earnings and foregone consumption are positively related to the length of sentences.

Punishments affect not only offenders but also other members of society. Aside from collection costs, fines paid by offenders are received as revenue by others. Most punishments, however, hurt other members as well as offenders: for example, imprisonment requires expenditures on guards, supervisory personnel, buildings, food, etc. Currently about $1 billion is being spent each year in the United States on probation, parole, and institutionalization alone, with the daily cost per case varying tremendously from a low of $0.38 for adults on probation to a high of $11.00 for juveniles in detention institutions (President's Commission, 1967*b*, pp. 193–94).

The total social cost of punishments is the cost to offenders plus the cost or minus the gain to others. Fines produce a gain to the latter that equals the cost to offenders, aside from collection costs, and so the social cost of fines is about zero, as befits a transfer payment. The social cost of probation, imprisonment, and other punishments, however, generally exceeds that to offenders, because others are also hurt. The derivation of optimality conditions in the next section is made more convenient if social costs are written in terms of offender costs as

$$f' \equiv bf, \tag{15}$$

where f' is the social cost and b is a coefficient that transforms f into f'. The size of b varies greatly between different kinds of punishments: $b \cong 0$ for fines, while $b > 1$ for torture, probation, parole, imprisonment, and most other punishments. It is especially large for juveniles in detention homes or for adults in prisons and is rather close to unity for torture or for adults on parole.

III. Optimality Conditions

The relevant parameters and behavioral functions have been introduced, and the stage is set for a discussion of social policy. If the aim simply were deterrence, the probability of conviction, p, could be raised close to 1, and punishments, f, could be made to exceed the gain: in this way the number of offenses, O, could be reduced almost at will. However, an increase in p increases the social cost of offenses through its effect on the cost of combatting offenses, C, as does an increase in f if $b > 0$ through the effect on the cost of punishments, bf. At relatively modest values of p and f, these effects might outweigh the social gain from increased deterrence. Similarly,

[21] In this respect, imprisonment is a special case of "waiting time" pricing that is also exemplified by queuing (see Becker, 1965, esp. pp. 515–16, and Kleinman, 1967).

Copyright © 2001. All Rights Reserved.

if the aim simply were to make "the punishment fit the crime," p could be set close to 1, and f could be equated to the harm imposed on the rest of society. Again, however, such a policy ignores the social cost of increases in p and f.

What is needed is a criterion that goes beyond catchy phrases and gives due weight to the damages from offenses, the costs of apprehending and convicting offenders, and the social cost of punishments. The social-welfare function of modern welfare economics is such a criterion, and one might assume that society has a function that measures the social loss from offenses. If

$$L = L(D, C, bf, O) \tag{16}$$

is the function measuring social loss, with presumably

$$\frac{\partial L}{\partial D} > 0, \quad \frac{\partial L}{\partial C} > 0, \quad \frac{\partial L}{\partial bf} > 0, \tag{17}$$

the aim would be to select values of f, C, and possibly b that minimize L.

It is more convenient and transparent, however, to develop the discussion at this point in terms of a less general formulation, namely, to assume that the loss function is identical with the total social loss in real income from offenses, convictions, and punishments, as in

$$L = D(O) + C(p, O) + bpfO. \tag{18}$$

The term $bpfO$ is the total social loss from punishments, since bf is the loss per offense punished and pO is the number of offenses punished (if there are a fairly large number of independent offenses). The variables directly subject to social control are the amounts spent in combatting offenses, C; the punishment per offense for those convicted, f; and the form of punishments, summarized by b. Once chosen, these variables, via the D, C, and O functions, indirectly determine p, O, D, and ultimately the loss L.

Analytical convenience suggests that p rather than C be considered a decision variable. Also, the coefficient b is assumed in this section to be a given constant greater than zero. Then p and f are the only decision variables, and their optimal values are found by differentiating L to find the two first-order optimality conditions,[22]

$$\frac{\partial L}{\partial f} = D'O_f + C'O_f + bpfO_f + bpO = 0 \tag{19}$$

and

$$\frac{\partial L}{\partial p} = D'O_p + C'O_p + C_p + bpfO_p + bfO = 0. \tag{20}$$

[22] The Mathematical Appendix discusses second-order conditions.

Copyright © 2001. All Rights Reseved.

If O_f and O_p are not equal to zero, one can divide through by them, and recombine terms, to get the more interesting expressions

$$D' + C' = -bpf\left(1 - \frac{1}{\varepsilon_f}\right) \tag{21}$$

and

$$D' + C' + C_p \frac{1}{O_p} = -bpf\left(1 - \frac{1}{\varepsilon_p}\right), \tag{22}$$

where

$$\varepsilon_f = -\frac{f}{O} O_f$$

and (23)

$$\varepsilon_p = -\frac{p}{O} O_p.$$

The term on the left side of each equation gives the marginal cost of increasing the number of offenses, O: in equation (21) through a reduction in f and in (22) through a reduction in p. Since $C' > 0$ and O is assumed to be in a region where $D' > 0$, the marginal cost of increasing O through f must be positive. A reduction in p partly reduces the cost of combatting offenses, and, therefore, the marginal cost of increasing O must be less when p rather than when f is reduced (see Fig. 1); the former could even be negative if C_p were sufficiently large. Average "revenue," given by $-bpf$, is negative, but marginal revenue, given by the right-hand side of

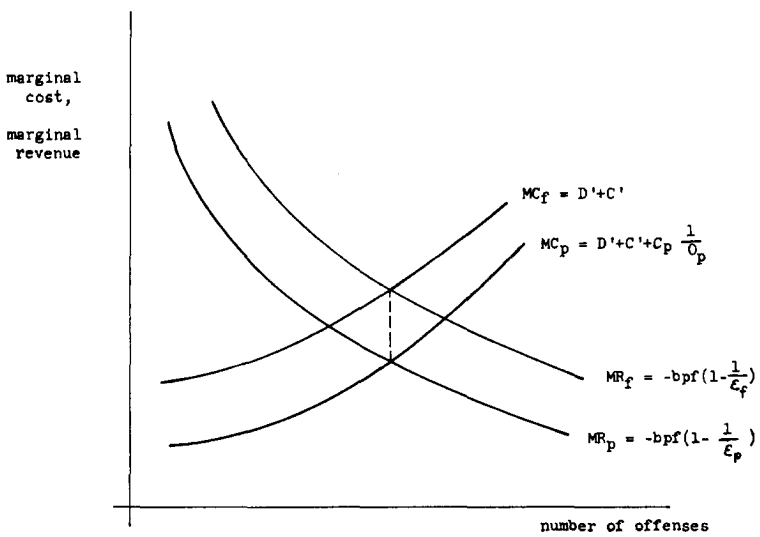

FIG. 1

Copyright © 2001. All Rights Reserved.

equations (21) and (22), is not necessarily negative and would be positive if the elasticities ε_p and ε_f were less than unity. Since the loss is minimized when marginal revenue equals marginal cost (see Fig. 1), the optimal value of ε_f must be less than unity, and that of ε_p could only exceed unity if C_p were sufficiently large. This is a reversal of the usual equilibrium condition for an income-maximizing firm, which is that the elasticity of demand must exceed unity, because in the usual case average revenue is assumed to be positive.[23]

Since the marginal cost of changing O through a change in p is less than that of changing O through f, the equilibrium marginal revenue from p must also be less than that from f. But equations (21) and (22) indicate that the marginal revenue from p can be less if, and only if, $\varepsilon_p > \varepsilon_f$. As pointed out earlier, however, this is precisely the condition indicating that offenders have preference for risk and thus that "crime does not pay." Consequently, the loss from offenses is minimized if p and f are selected from those regions where offenders are, on balance, risk preferrers. Although only the attitudes offenders have toward risk can directly determine whether "crime pays," rational public policy indirectly insures that "crime does not pay" through its choice of p and f.[24]

I indicated earlier that the actual p's and f's for major felonies in the United States generally seem to be in regions where the effect (measured by elasticity) of p on offenses exceeds that of f, that is, where offenders are risk preferrers and "crime does not pay" (Smigel, 1965; Ehrlich, 1967). Moreover, both elasticities are generally less than unity. In both respects, therefore, actual public policy is consistent with the implications of the optimality analysis.

If the supply of offenses depended only on pf—offenders were risk neutral—a reduction in p "compensated" by an equal percentage increase in f would leave unchanged pf, O, $D(O)$, and $bpfO$ but would reduce the loss, because the costs of apprehension and conviction would be lowered by the reduction in p. The loss would be minimized, therefore, by lowering p arbitrarily close to zero and raising f sufficiently high so that the product pf would induce the optimal number of offenses.[25] A fortiori, if offenders

[23] Thus if $b < 0$, average revenue would be positive and the optimal value of ε_f would be greater than 1, and that of ε_p could be less than 1 only if C_p were sufficiently large.

[24] If $b < 0$, the optimality condition is that $\varepsilon_p < \varepsilon_f$, or that offenders are risk avoiders. Optimal social policy would then be to select p and f in regions where "crime does pay."

[25] Since $\varepsilon_f = \varepsilon_p = \varepsilon$ if O depends only on pf, and $C = 0$ if $p = 0$, the two equilibrium conditions given by eqs. (21) and (22) reduce to the single condition

$$D' = -bpf\left(1 - \frac{1}{\varepsilon}\right).$$

From this condition and the relation $O = O(pf)$, the equilibrium values of O and pf could be determined.

Copyright © 2001. All Rights Reseved.

were risk avoiders, the loss would be minimized by setting p arbitrarily close to zero, for a "compensated" reduction in p reduces not only C but also O and thus D and $bpfO$.[26]

There was a tendency during the eighteenth and nineteenth centuries in Anglo-Saxon countries, and even today in many Communist and underdeveloped countries, to punish those convicted of criminal offenses rather severely, at the same time that the probability of capture and conviction was set at rather low values.[27] A promising explanation of this tendency is that an increased probability of conviction obviously absorbs public and private resources in the form of more policemen, judges, juries, and so forth. Consequently, a "compensated" reduction in this probability obviously reduces expenditures on combatting crime, and, since the expected punishment is unchanged, there is no "obvious" offsetting increase in either the amount of damages or the cost of punishments. The result can easily be continuous political pressure to keep police and other expenditures relatively low and to compensate by meting out strong punishments to those convicted.

Of course, if offenders are risk preferrers, the loss in income from offenses is generally minimized by selecting positive and finite values of p and f, even though there is no "obvious" offset to a compensated reduction in p. One possible offset already hinted at in footnote 27 is that judges or juries may be unwilling to convict offenders if punishments are set very high. Formally, this means that the cost of apprehension and conviction, C, would depend not only on p and O but also on f.[28] If C were more responsive to f than p, at least in some regions,[29] the loss in income could be minimized at finite values of p and f even if offenders were risk avoiders. For then a compensated reduction in p could raise, rather than lower, C and thus contribute to an increase in the loss.

Risk avoidance might also be consistent with optimal behavior if the loss function were not simply equal to the reduction in income. For example, suppose that the loss were increased by an increase in the ex post "price discrimination" between offenses that are not and those that are cleared by punishment. Then a "compensated" reduction in p would

[26] If $b < 0$, the optimal solution is p about zero and f arbitrarily high if offenders are either risk neutral or risk preferrers.

[27] For a discussion of English criminal law in the eighteenth and nineteenth centuries, see Radzinowicz (1948, Vol. I). Punishments were severe then, even though the death penalty, while legislated, was seldom implemented for less serious criminal offenses.

Recently South Vietnam executed a prominent businessman allegedly for "speculative" dealings in rice, while in recent years a number of persons in the Soviet Union have either been executed or given severe prison sentences for economic crimes.

[28] I owe the emphasis on this point to Evsey Domar.

[29] This is probably more likely for higher values of f and lower values of p.

Copyright © 2001 - All Rights Reserved

increase the "price discrimination," and the increased loss from this could more than offset the reductions in C, D, and $bpfO$.[30]

IV. Shifts in the Behavioral Relations

This section analyzes the effects of shifts in the basic behavioral relations—the damage, cost, and supply-of-offenses functions—on the optimal values of p and f. Since rigorous proofs can be found in the Mathematical Appendix, here the implications are stressed, and only intuitive proofs are given. The results are used to explain, among other things, why more damaging offenses are punished more severely and more impulsive offenders less severely.

An increase in the marginal damages from a given number of offenses, D', increases the marginal cost of changing offenses by a change in either p or f (see Fig. 2*a* and *b*). The optimal number of offenses would necessarily decrease, because the optimal values of both p and f would increase. In this case (and, as shortly seen, in several others), the optimal values of p and f move in the same, rather than in opposite, directions.[31]

An interesting application of these conclusions is to different kinds of offenses. Although there are few objective measures of the damages done

[30] If p is the probability that an offense would be cleared with the punishment f, then $1 - p$ is the probability of no punishment. The expected punishment would be $\mu = pf$, the variance $\sigma^2 = p(1 - p)f^2$, and the coefficient of variation

$$v = \frac{\sigma}{\mu} = \sqrt{\frac{1-p}{p}};$$

v increases monotonically from a low of zero when $p = 1$ to an infinitely high value when $p = 0$.

If the loss function equaled

$$L' = L + \psi(v), \qquad \psi' > 0,$$

the optimality conditions would become

$$D' + C' = -bpf\left(1 - \frac{1}{\varepsilon_f}\right) \tag{21}$$

and

$$D' + C' + C_p \frac{1}{O_p} + \psi' \frac{dv}{dp}\frac{1}{O_p} = -bpf\left(1 - \frac{1}{\varepsilon_p}\right). \tag{22}$$

Since the term $\psi'(dv/dp)(1/O_p)$ is positive, it could more than offset the negative term $C_p(1/O_p)$.

[31] I stress this primarily because of Bentham's famous and seemingly plausible dictum that "the more deficient in certainty a punishment is, the severer it should be" (1931, chap. ii of section entitled "Of Punishment," second rule). The dictum would be correct if p (or f) were exogenously determined and if L were minimized with respect to f (or p) alone, for then the optimal value of f (or p) would be inversely related to the given value of p (or f) (see the Mathematical Appendix). If, however, L is minimized with respect to both, then frequently they move in the same direction.

Copyright © 2001. All Rights Reseved.

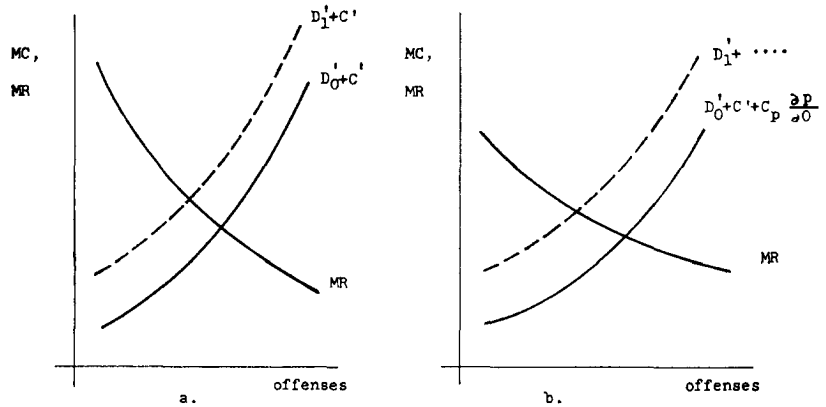

FIG. 2

by most offenses, it does not take much imagination to conclude that offenses like murder or rape generally do more damage than petty larceny or auto theft. If the other components of the loss in income were the same, the optimal probability of apprehension and conviction and the punishment when convicted would be greater for the more serious offenses.

Table 2 presents some evidence on the actual probabilities and punishments in the United States for seven felonies. The punishments are simply the average prison sentences served, while the probabilities are ratios of the estimated number of convictions to the estimated number of offenses and unquestionably contain a large error (see the discussions in Smigel, 1965, and Ehrlich, 1967). If other components of the loss function are ignored, and if actual and optimal probabilities and punishments are positively related, one should find that the more serious felonies have higher probabilities and longer prison terms. And one does: in the table, which lists the felonies in decreasing order of presumed seriousness, both the actual probabilities and the prison terms are positively related to seriousness.

Since an increase in the marginal cost of apprehension and conviction for a given number of offenses, C', has identical effects as an increase in marginal damages, it must also reduce the optimal number of offenses and increase the optimal values of p and f. On the other hand, an increase in the other component of the cost of apprehension and conviction, C_p, has no direct effect on the marginal cost of changing offenses with f and *reduces* the cost of changing offenses with p (see Fig. 3). It therefore reduces the optimal value of p and only partially compensates with an increase in f, so that the optimal number of offenses increases. Accordingly, an increase in both C' and C_p must increase the optimal f but can

Copyright © 2001. All Rights Reserved.

TABLE 2

PROBABILITY OF CONVICTION AND AVERAGE PRISON TERM FOR SEVERAL MAJOR FELONIES, 1960

	Murder and Non-negligent Manslaughter	Forcible Rape	Robbery	Aggravated Assault	Burglary	Larceny	Auto Theft	All These Felonies Combined
1. Average time served (months) before first release:								
a) Federal civil institutions	111.0	63.6	56.1	27.1	26.2	16.2	20.6	18.8
b) State institutions	121.4	44.8	42.4	25.0	24.6	19.8	21.3	28.4
2. Probabilities of apprehension and conviction (per cent):								
a) Those found guilty of offenses known	57.9	37.7	25.1	27.3	13.0	10.7	13.7	15.1
b) Those found guilty of offenses charged	40.7	26.9	17.8	16.1	10.2	9.8	11.5	15.0
c) Those entering federal and state prisons (excludes many juveniles)	39.8	22.7	8.4	3.0	2.4	2.2	2.1	2.8

Source: 1, Bureau of Prisons (1960, Table 3); 2 (*a*) and (*b*), Federal Bureau of Investigation (1960, Table 10); 2 (*c*), Federal Bureau of Investigation (1961, Table 2), Bureau of Prisons (n.d., Table A1; 1961, Table 8).

Copyright © 2001. All Rights Reseved.

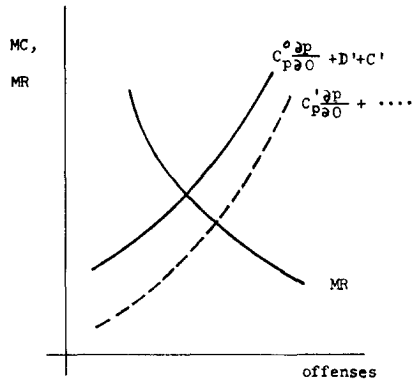

FIG. 3

either increase or decrease the optimal p and optimal number of offenses, depending on the relative importance of the changes in C' and C_p.

The cost of apprehending and convicting offenders is affected by a variety of forces. An increase in the salaries of policemen increases both C' and C_p, while improved police technology in the form of fingerprinting, ballistic techniques, computer control, and chemical analysis, or police and court "reform" with an emphasis on professionalism and merit, would tend to reduce both, not necessarily by the same extent. Our analysis implies, therefore, that although an improvement in technology and reform may or may not increase the optimal p and reduce the optimal number of offenses, it does reduce the optimal f and thus the need to rely on severe punishments for those convicted. Possibly this explains why the secular improvement in police technology and reform has gone hand in hand with a secular decline in punishments.

C_p, and to a lesser extent C', differ significantly between different kinds of offenses. It is easier, for example, to solve a rape or armed robbery than a burglary or auto theft, because the evidence of personal identification is often available in the former and not in the latter offenses.[32] This might tempt one to argue that the p's decline significantly as one moves across Table 2 (left to right) primarily because the C_p's are significantly lower for the "personal" felonies listed to the left than for the "impersonal" felonies listed to the right. But this implies that the f's would increase as one moved across the table, which is patently false. Consequently, the positive correlation between p, f, and the severity of offenses observed in

[32] "If a suspect is neither known to the victim nor arrested at the scene of the crime, the chances of ever arresting him are very slim" (President's Commission, 1967*e*, p. 8). This conclusion is based on a study of crimes in parts of Los Angeles during January, 1966.

Copyright © 2001 All Rights Reserved

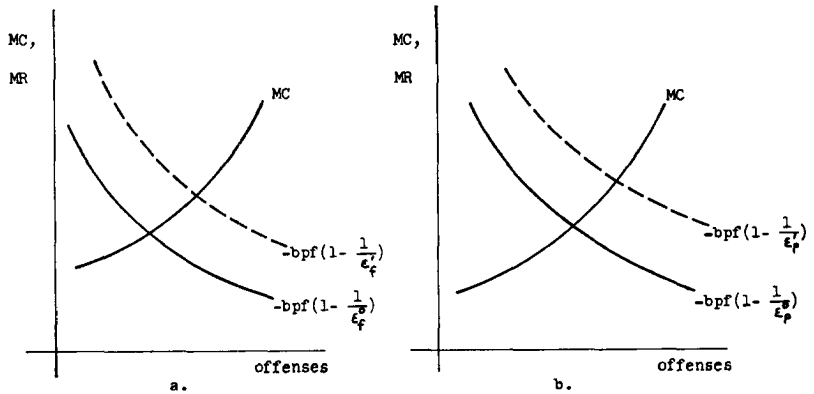

FIG. 4

the table cannot be explained by a negative correlation between C_p (or C') and severity.

If $b > 0$, a reduction in the elasticity of offenses with respect to f increases the marginal revenue of changing offenses by changing f (see Fig. 4*a*). The result is an increase in the optimal number of offenses and a decrease in the optimal f that is partially compensated by an increase in the optimal p. Similarly, a reduction in the elasticity of offenses with respect to p also increases the optimal number of offenses (see Fig. 4*b*), decreases the optimal p, and partially compensates by an increase in f. An equal percentage reduction in both elasticities a fortiori increases the optimal number of offenses and also tends to reduce both p and f. If $b = 0$, both marginal revenue functions lie along the horizontal axis, and changes in these elasticities have no effect on the optimal values of p and f.

The income of a firm would usually be larger if it could separate, at little cost, its total market into submarkets that have substantially different elasticities of demand: higher prices would be charged in the submarkets having lower elasticities. Similarly, if the total "market" for offenses could be separated into submarkets that differ significantly in the elasticities of supply of offenses, the results above imply that if $b > 0$ the total loss would be reduced by "charging" *lower* "prices"—that is, lower p's and f's—in markets with *lower* elasticities.

Sometimes it is possible to separate persons committing the same offense into groups that have different responses to punishments. For example, unpremeditated murderers or robbers are supposed to act impulsively and, therefore, to be relatively unresponsive to the size of punishments; likewise, the insane or the young are probably less affected

Copyright © 2001. All Rights Reseved.

than other offenders by future consequences and, therefore,[33] probably less deterred by increases in the probability of conviction or in the punishment when convicted. The trend during the twentieth century toward relatively smaller prison terms and greater use of probation and therapy for such groups and, more generally, the trend away from the doctrine of "a given punishment for a given crime" is apparently at least broadly consistent with the implications of the optimality analysis.

An increase in b increases the marginal revenue from changing the number of offenses by changing p or f and thereby increases the optimal number of offenses, reduces the optimal value of f, and increases the optimal value of p. Some evidence presented in Section II indicates that b is especially large for juveniles in detention homes or adults in prison and is small for fines or adults on parole. The analysis implies, therefore, that other things the same, the optimal f's would be smaller and the optimal p's larger if punishment were by one of the former rather than one of the latter methods.

V. Fines

A. Welfare Theorems and Transferable Pricing

The usual optimality conditions in welfare economics depend only on the levels and not on the slopes of marginal cost and average revenue functions, as in the well-known condition that marginal costs equal prices. The social loss from offenses was explicitly introduced as an application of the approach used in welfare economics, and yet slopes as incorporated into elasticities of supply do significantly affect the optimality conditions. Why this difference? The primary explanation would appear to be that it is almost always implicitly assumed that prices paid by consumers are fully transferred to firms and governments, so that there is no social loss from payment.

If there were no social loss from punishments, as with fines, b would equal zero, and the elasticity of supply would drop out of the optimality condition given by equation (21).[34] If $b > 0$, as with imprisonment, some of the payment "by" offenders would not be received by the rest of society, and a net social loss would result. The elasticity of the supply of offenses then becomes an important determinant of the optimality conditions, because it determines the change in social costs caused by a change in punishments.

[33] But see Becker (1962) for an analysis indicating that impulsive and other "irrational" persons may be as deterred from purchasing a commodity whose price has risen as more "rational" persons.

[34] It remains in eq. (22), through the slope O_p, because ordinarily prices do not affect marginal costs, while they do here through the influence of p on C.

Copyright © 2001 All Rights Reserved

Although transferable monetary pricing is the most common kind today, the other is not unimportant, especially in underdeveloped and Communist countries. Examples in addition to imprisonment and many other punishments are the draft, payments in kind, and queues and other waiting-time forms of rationing that result from legal restrictions on pricing (see Becker, 1965) and from random variations in demand and supply conditions. It is interesting, and deserves further exploration, that the optimality conditions are so significantly affected by a change in the assumptions about the transferability of pricing.

B. Optimality Conditions

If $b = 0$, say, because punishment was by fine, and if the cost of apprehending and convicting offenders were also zero, the two optimality conditions (21) and (22) would reduce to the same simple condition

$$D'(O) = 0. \tag{24}$$

Economists generally conclude that activities causing "external" harm, such as factories that pollute the air or lumber operations that strip the land, should be taxed or otherwise restricted in level until the marginal external harm equalled the marginal private gain, that is, until marginal net damages equalled zero, which is what equation (24) says. If marginal harm always exceeded marginal gain, the optimum level would be presumed to be zero, and that would also be the implication of (24) when suitable inequality conditions were brought in. In other words, if the costs of apprehending, convicting, and punishing offenders were nil and if each offense caused more external harm than private gain, the social loss from offenses would be minimized by setting punishments high enough to eliminate all offenses. Minimizing the social loss would become identical with the criterion of minimizing crime by setting penalties sufficiently high.[35]

Equation (24) determines the optimal number of offenses, $\hat{O}$, and the fine and probability of conviction must be set at levels that induce offenders to commit just $\hat{O}$ offenses. If the economists' usual theory of choice is applied to illegal activities (see Sec. II), the marginal value of these penalties has to equal the marginal private gain:

$$V = G'(\hat{O}), \tag{25}$$

where $G'(\hat{O})$ is the marginal private gain at $\hat{O}$ and V is the monetary value of the marginal penalties. Since by equations (3) and (24), $D'(\hat{O}) = H'(\hat{O}) - G'(\hat{O}) = 0$, one has by substitution in (25)

$$V = H'(\hat{O}). \tag{26}$$

[35] "The evil of the punishment must be made to exceed the advantage of the offense" (Bentham, 1931, first rule).

Copyright © 2001. All Rights Reseved.

The monetary value of the penalties would equal the marginal harm caused by offenses.

Since the cost of apprehension and conviction is assumed equal to zero, the probability of apprehension and conviction could be set equal to unity without cost. The monetary value of penalties would then simply equal the fines imposed, and equation (26) would become

$$f = H'(\hat{O}). \tag{27}$$

Since fines are paid by offenders to the rest of society, a fine determined by (27) would exactly compensate the latter for the marginal harm suffered, and the criterion of minimizing the social loss would be identical, at the margin, with the criterion of compensating "victims."[36] If the harm to victims always exceeded the gain to offenders, both criteria would reduce in turn to eliminating all offenses.

If the cost of apprehension and conviction were not zero, the optimality condition would have to incorporate marginal costs as well as marginal damages and would become, if the probability of conviction were still assumed to equal unity,

$$D'(\hat{O}) + C'(\hat{O}, 1) = 0. \tag{28}$$

Since $C' > 0$, (28) requires that $D' < 0$ or that the marginal private gain exceed the marginal external harm, which generally means a smaller number of offenses than when $D' = 0$.[37] It is easy to show that equation (28) would be satisfied if the fine equalled the sum of marginal harm and marginal costs:

$$f = H'(\hat{O}) + C'(\hat{O}, 1).^{38} \tag{29}$$

In other words, offenders have to compensate for the cost of catching them as well as for the harm they directly do, which is a natural generalization of the usual externality analysis.

The optimality condition

$$D'(\hat{O}) + C'(\hat{O}, \hat{p}) + C_p(\hat{O}, \hat{p})\frac{1}{O_p} = 0 \tag{30}$$

would replace equation (28) if the fine rather than the probability of

[36] By "victims" is meant the rest of society and not just the persons actually harmed.

[37] This result can also be derived as a special case of the results in the Mathematical Appendix on the effects of increases in C'.

[38] Since equilibrium requires that $f = G'(\hat{O})$, and since from (28)

$$D'(\hat{O}) = H'(\hat{O}) - G'(\hat{O}) = -C'(\hat{O}, 1),$$

then (29) follows directly by substitution.

Copyright © 2001. All Rights Reserved.

conviction were fixed. Equation (30) would usually imply that $D'(\hat{O}) > 0$,[39] and thus that the number of offenses would exceed the optimal number when costs were zero. Whether costs of apprehension and conviction increase or decrease the optimal number of offenses largely depends, therefore, on whether penalties are changed by a change in the fine or in the probability of conviction. Of course, if both are subject to control, the optimal probability of conviction would be arbitrarily close to zero, unless the social loss function differed from equation (18) (see the discussion in Sec. III).

C. The Case for Fines

Just as the probability of conviction and the severity of punishment are subject to control by society, so too is the form of punishment: legislation usually specifies whether an offense is punishable by fines, probation, institutionalization, or some combination. Is it merely an accident, or have optimality considerations determined that today, in most countries, fines are the predominant form of punishment, with institutionalization reserved for the more serious offenses? This section presents several arguments which imply that social welfare is increased if fines are used *whenever feasible.*

In the first place, probation and institutionalization use up social resources, and fines do not, since the latter are basically just transfer payments, while the former use resources in the form of guards, supervisory personnel, probation officers, and the offenders' own time.[40] Table 1 indicates that the cost is not minor either: in the United States in 1965, about $1 billion was spent on "correction," and this estimate excludes, of course, the value of the loss in offenders' time.[41]

[39] That is, if, as seems plausible,

$$\frac{dC}{dp} = C'\frac{\partial O}{\partial p} + C_p > 0,$$

then

$$C' + C_p \frac{1}{\partial O/\partial p} < 0,$$

and

$$D'(\hat{O}) = -\left(C' + C_p \frac{1}{\partial O/\partial p}\right) > 0.$$

[40] Several early writers on criminology recognized this advantage of fines. For example, "Pecuniary punishments are highly economical, since all the evil felt by him who pays turns into an advantage for him who receives" (Bentham, 1931, chap. vi), and "Imprisonment would have been regarded in these old times [*ca.* tenth century] as a useless punishment; it does not satisfy revenge, it keeps the criminal idle, and do what we may, *it is costly*" (Pollock and Maitland, 1952, p. 516; my italics).

[41] On the other hand, some transfer payments in the form of food, clothing, and shelter are included.

Copyright © 2001. All Rights Reseved.

Moreover, the determination of the optimal number of offenses and severity of punishments is somewhat simplified by the use of fines. A wise use of fines requires knowledge of marginal gains and harm and of marginal apprehension and conviction costs; admittedly, such knowledge is not easily acquired. A wise use of imprisonment and other punishments must know this too, however, and, in addition, must know about the elasticities of response of offenses to changes in punishments. As the bitter controversies over the abolition of capital punishment suggest, it has been difficult to learn about these elasticities.

I suggested earlier that premeditation, sanity, and age can enter into the determination of punishments as proxies for the elasticities of response. These characteristics may not have to be considered in levying fines, because the optimal fines, as determined, say, by equations (27) or (29), do not depend on elasticities. Perhaps this partly explains why economists discussing externalities almost never mention motivation or intent, while sociologists and lawyers discussing criminal behavior invariably do. The former assume that punishment is by a monetary tax or fine, while the latter assume that non-monetary punishments are used.

Fines provide compensation to victims, and optimal fines at the margin fully compensate victims and restore the status quo ante, so that they are no worse off than if offenses were not committed.[42] Not only do other punishments fail to compensate, but they also require "victims" to spend additional resources in carrying out the punishment. It is not surprising, therefore, that the anger and fear felt toward ex-convicts who in fact have *not* "paid their debt to society" have resulted in additional punishments,[43] including legal restrictions on their political and economic opportunities[44] and informal restrictions on their social acceptance. Moreover, the absence of compensation encourages efforts to change and otherwise "rehabilitate" offenders through psychiatric counseling, therapy, and other programs. Since fines do compensate and do not create much additional cost, anger toward and fear of appropriately fined persons do not easily develop. As a result, additional punishments are not usually levied against "ex-finees," nor are strong efforts made to "rehabilitate" them.

One argument made against fines is that they are immoral because, in effect, they permit offenses to be bought for a price in the same way that

[42] Bentham recognized this and said, "To furnish an indemnity to the injured party is another useful quality in a punishment. It is a means of accomplishing two objects at once—punishing an offense and repairing it: removing the evil of the first order, and putting a stop to alarm. This is a characteristic advantage of pecuniary punishments" (1931, chap. vi).

[43] In the same way, the guilt felt by society in using the draft, a forced transfer *to* society, has led to additional payments to veterans in the form of education benefits, bonuses, hospitalization rights, etc.

[44] See Sutherland (1960, pp. 267–68) for a list of some of these.

Copyright © 2001. All Rights Reserved.

bread or other goods are bought for a price.[45] A fine *can* be considered the price of an offense, but so too can any other form of punishment; for example, the "price" of stealing a car might be six months in jail. The only difference is in the units of measurement: fines are prices measured in monetary units, imprisonments are prices measured in time units, etc. If anything, monetary units are to be preferred here as they are generally preferred in pricing and accounting.

Optimal fines determined from equation (29) depend only on the marginal harm and cost and not at all on the economic positions of offenders. This has been criticized as unfair, and fines proportional to the incomes of offenders have been suggested.[46] If the goal is to minimize the social loss in income from offenses, and not to take vengeance or to inflict harm on offenders, then fines should depend on the total harm done by offenders, and not directly on their income, race, sex, etc. In the same way, the monetary value of optimal prison sentences and other punishments depends on the harm, costs, and elasticities of response, but not directly on an offender's income. Indeed, if the monetary value of the punishment by, say, imprisonment were independent of income, the length of the sentence would be *inversely* related to income, because the value placed on a given sentence is positively related to income.

We might detour briefly to point out some interesting implications for the probability of conviction of the fact that the monetary value of a given fine is obviously the same for all offenders, while the monetary equivalent or "value" of a given prison sentence or probation period is generally positively related to an offender's income. The discussion in Section II suggested that actual probabilities of conviction are not fixed to all offenders but usually vary with their age, sex, race, and, in particular, income. Offenders with higher earnings have an incentive to spend more on planning their offenses, on good lawyers, on legal appeals, and even on bribery to reduce the probability of apprehension and conviction for offenses punishable by, say, a given prison term, because the cost to them of conviction is relatively large compared to the cost of these expenditures.

[45] The very early English law relied heavily on monetary fines, even for murder, and it has been said that "every kind of blow or wound given to every kind of person had its price, and much of the jurisprudence of the time must have consisted of a knowledge of these preappointed prices" (Pollock and Maitland, 1952, p. 451).

The same idea was put amusingly in a recent *Mutt and Jeff* cartoon which showed a police car carrying a sign that read: "Speed limit 30 M per H—$5 fine every mile over speed limit—pick out speed you can afford."

[46] For example, Bentham said, "A pecuniary punishment, if the sum is fixed, is in the highest degree unequal. . . . Fines have been determined without regard to the profit of the offense, to its evil, or to the wealth of the offender. . . . Pecuniary punishments should always be regulated by the fortune of the offender. The relative amount of the fine should be fixed, not its absolute amount; for such an offense, such a part of the offender's fortune" (1931, chap. ix). Note that optimal fines, as determined by eq. (29), do depend on "the profit of the offense" and on "its evil."

Copyright © 2001. All Rights Reseved.

Similarly, however, poorer offenders have an incentive to use more of their time in planning their offenses, in court appearances, and the like to reduce the probability of conviction for offenses punishable by a given fine, because the cost to them of conviction is relatively large compared to the value of their time.[47] The implication is that the probability of conviction would be systematically related to the earnings of offenders: negatively for offenses punishable by imprisonment and positively for those punishable by fines. Although a negative relation for felonies and other offenses punishable by imprisonment has been frequently observed and deplored (see President's Commission, 1967*c*, pp. 139–53), I do not know of any studies of the relation for fines or of any recognition that the observed negative relation may be more a consequence of the nature of the punishment than of the influence of wealth.

Another argument made against fines is that certain crimes, like murder or rape, are so heinous that no amount of money could compensate for the harm inflicted. This argument has obvious merit and is a special case of the more general principle that fines cannot be relied on exclusively whenever the harm exceeds the resources of offenders. For then victims could not be fully compensated by offenders, and fines would have to be supplemented with prison terms or other punishments in order to discourage offenses optimally. This explains why imprisonments, probation, and parole are major punishments for the more serious felonies; considerable harm is inflicted, and felonious offenders lack sufficient resources to compensate. Since fines are preferable, it also suggests the need for a flexible system of instalment fines to enable offenders to pay fines more readily and thus avoid other punishments.

This analysis implies that if some offenders could pay the fine for a given offense and others could not,[48] the former should be punished solely by fine and the latter partly by other methods. In essence, therefore, these methods become a vehicle for punishing "debtors" to society. Before the cry is raised that the system is unfair, especially to poor offenders, consider the following.

Those punished would be debtors in "transactions" that were never agreed to by their "creditors," not in voluntary transactions, such as loans,[49] for which suitable precautions could be taken in advance by creditors. Moreover, punishment in any economic system based on

[47] Note that the incentive to use time to reduce the probability of a given prison sentence is unrelated to earnings, because the punishment is fixed in time, not monetary, units; likewise, the incentive to use money to reduce the probability of a given fine is also unrelated to earnings, because the punishment is fixed in monetary, not time, units.

[48] In one study, about half of those convicted of misdemeanors could not pay the fines (see President's Commission, 1967*c*, p. 148).

[49] The "debtor prisons" of earlier centuries generally housed persons who could not repay loans.

Copyright © 2001. All Rights Reserved.

voluntary market transactions inevitably must distinguish between such "debtors" and others. If a rich man purchases a car and a poor man steals one, the former is congratulated, while the latter is often sent to prison when apprehended. Yet the rich man's purchase is equivalent to a "theft" subsequently compensated by a "fine" equal to the price of the car, while the poor man, in effect, goes to prison because he cannot pay this "fine."

Whether a punishment like imprisonment in lieu of a full fine for offenders lacking sufficient resources is "fair" depends, of course, on the length of the prison term compared to the fine.[50] For example, a prison term of one week in lieu of a $10,000 fine would, if anything, be "unfair" to wealthy offenders paying the fine. Since imprisonment is a more costly punishment to society than fines, the loss from offenses would be reduced by a policy of leniency toward persons who are imprisoned because they cannot pay fines. Consequently, optimal prison terms for "debtors" would not be "unfair" to them in the sense that the monetary equivalent to them of the prison terms would be less than the value of optimal fines, which in turn would equal the harm caused or the "debt."[51]

It appears, however, that "debtors" are often imprisoned at rates of exchange with fines that place a low value on time in prison. Although I have not seen systematic evidence on the different punishments actually offered convicted offenders, and the choices they made, many statutes in

[50] Yet without any discussion of the actual alternatives offered, the statement is made that "the money judgment assessed the punitive damages defendant hardly seems comparable in effect to the criminal sanctions of death, imprisonment, and stigmatization" ("Criminal Safeguards . . .," 1967).

[51] A formal proof is straightforward if for simplicity the probability of conviction is taken as equal to unity. For then the sole optimality condition is

$$D' + C' = -bf\left(1 - \frac{1}{\varepsilon_f}\right). \tag{1'}$$

Since $D' = H' - G'$, by substitution one has

$$G' = H' + C' + bf\left(1 - \frac{1}{\varepsilon_f}\right), \tag{2'}$$

and since equilibrium requires that $G' = f$,

$$f = H' + C' + bf\left(1 - \frac{1}{\varepsilon_f}\right), \tag{3'}$$

or

$$f = \frac{H' + C'}{1 - b(1 - 1/\varepsilon_f)}. \tag{4'}$$

If $b > 0$, $\varepsilon_f < 1$ (see Sec. III), and hence by eq. (4′),

$$f < H' + C', \tag{5'}$$

where the term on the right is the full marginal harm. If p as well as f is free to vary, the analysis becomes more complicated, but the conclusion about the relative monetary values of optimal imprisonments and fines remains the same (see the Mathematical Appendix).

Copyright © 2001. All Rights Reseved.

the United States do permit fines and imprisonment that place a low value on time in prison. For example, in New York State, Class A Misdemeanors can be punished by a prison term as long as one year or a fine no larger than $1,000 and Class B Misdemeanors, by a term as long as three months or a fine no larger than $500 (*Laws of New York*, 1965, chap. 1030, Arts. 70 and 80).[52] According to my analysis, these statutes permit excessive prison sentences relative to the fines, which may explain why imprisonment in lieu of fines is considered unfair to poor offenders, who often must "choose" the prison alternative.

D. Compensation and the Criminal Law

Actual criminal proceedings in the United States appear to seek a mixture of deterrence, compensation, and vengeance. I have already indicated that these goals are somewhat contradictory and cannot generally be simultaneously achieved; for example, if punishment were by fine, minimizing the social loss from offenses would be equivalent to compensating "victims" fully, and deterrence or vengeance could only be partially pursued. Therefore, if the case for fines were accepted, and punishment by optimal fines became the norm, the traditional approach to criminal law would have to be significantly modified.

First and foremost, the primary aim of all legal proceedings would become the same: not punishment or deterrence, but simply the assessment of the "harm" done by defendants. Much of traditional criminal law would become a branch of the law of torts,[53] say "social torts," in which the public would collectively sue for "public" harm. A "criminal" action would be defined fundamentally not by the nature of the action[54] but by the inability of a person to compensate for the "harm" that he caused. Thus an action would be "criminal" precisely because it results in uncompensated "harm" to others. Criminal law would cover all such actions, while tort law would cover all other (civil) actions.

As a practical example of the fundamental changes that would be wrought, consider the antitrust field. Inspired in part by the economist's classic demonstration that monopolies distort the allocation of resources and reduce economic welfare, the United States has outlawed conspiracies

[52] "Violations," however, can only be punished by prison terms as long as fifteen days or fines no larger than $250. Since these are maximum punishments, the actual ones imposed by the courts can, and often are, considerably less. Note, too, that the courts can punish by imprisonment, by fine, or by *both* (*Laws of New York*, 1965, chap. 1030, Art. 60).

[53] "The cardinal principle of damages in Anglo-American law [of torts] is that of *compensation* for the injury caused to plaintiff by defendant's breach of duty" (Harper and James, 1956, p. 1299).

[54] Of course, many traditional criminal actions like murder or rape would still usually be criminal under this approach too.

Copyright © 2001 All Rights Reserved

and other constraints of trade. In practice, defendants are often simply required to cease the objectionable activity, although sometimes they are also fined, become subject to damage suits, or are jailed.

If compensation were stressed, the main purpose of legal proceedings would be to levy fines equal to[55] the harm inflicted on society by constraints of trade. There would be no point to cease and desist orders, imprisonment, ridicule, or dissolution of companies. If the economist's theory about monopoly is correct, and if optimal fines were levied, firms would automatically cease any constraints of trade, because the gain to them would be less than the harm they cause and thus less than the fines expected. On the other hand, if Schumpeter and other critics are correct, and certain constraints of trade raise the level of economic welfare, fines could fully compensate society for the harm done, and yet some constraints would not cease, because the gain to participants would exceed the harm to others.[56]

One unexpected advantage, therefore, from stressing compensation and fines rather than punishment and deterrence is that the validity of the classical position need not be judged a priori. If valid, compensating fines would discourage all constraints of trade and would achieve the classical aims. If not, such fines would permit the socially desirable constraints to continue and, at the same time, would compensate society for the harm done.

Of course, as participants in triple-damage suits are well aware, the harm done is not easily measured, and serious mistakes would be inevitable. However, it is also extremely difficult to measure the harm in many civil suits,[57] yet these continue to function, probably reasonably well on the whole. Moreover, as experience accumulated, the margin of error would decline, and rules of thumb would develop. Finally, one must realize that difficult judgments are also required by the present antitrust policy, such as deciding that certain industries are "workably" competitive or that certain mergers reduce competition. An emphasis on fines and compensation would at least help avoid irrelevant issues by focusing attention on the information most needed for intelligent social policy.

[55] Actually, fines should exceed the harm done if the probability of conviction were less than unity. The possibility of avoiding conviction is the intellectual justification for punitive, such as triple, damages against those convicted.

[56] The classical view is that $D'(M)$ always is greater than zero, where M measures the different constraints of trade and D' measures the marginal damage; the critic's view is that for some M, $D'(M) < 0$. It has been shown above that if D' always is greater than zero, compensating fines would discourage all offenses, in this case constraints of trade, while if D' sometimes is less than zero, some offenses would remain (unless $C'[M]$, the marginal cost of detecting and convicting offenders, were sufficiently large relative to D').

[57] Harper and James said, "Sometimes [compensation] can be accomplished with a fair degree of accuracy. But obviously it cannot be done in anything but a figurative and essentially speculative way for many of the consequences of personal injury. Yet it is the aim of the law to attain at least a rough correspondence between the amount awarded as damages and the extent of the suffering" (1956, p. 1301).

Copyright © 2001. All Rights Reseved.

VI. Private Expenditures against Crime

A variety of private as well as public actions also attempt to reduce the number and incidence of crimes: guards, doormen, and accountants are employed, locks and alarms installed, insurance coverage extended, parks and neighborhoods avoided, taxis used in place of walking or subways, and so on. Table 1 lists close to $2 billion of such expenditures in 1965, and this undoubtedly is a gross underestimate of the total. The need for private action is especially great in highly interdependent modern economies, where frequently a person must trust his resources, including his person, to the "care" of employees, employers, customers, or sellers.

If each person tries to minimize his expected loss in income from crimes, optimal private decisions can be easily derived from the previous discussion of optimal public ones. For each person there is a loss function similar to that given by equation (18):

$$L_j = H_j(O_j) + C_j(p_j, O_j, C, C_k) + b_j p_j f_j O_j. \tag{31}$$

The term H_j represents the harm to j from the O_j offenses committed against j, while C_j represents his cost of achieving a probability of conviction of p_j for offenses committed against him. Note that C_j not only is positively related to O_j but also is negatively related to C, public expenditures on crime, and to C_k, the set of private expenditures by other persons.[58]

The term $b_j p_j f_j O_j$ measures the expected[59] loss to j from punishment of offenders committing any of the O_j. Whereas most punishments result in a net loss to society as a whole, they often produce a gain for the actual victims. For example, punishment by fines given to the actual victims is just a transfer payment for society but is a clear gain to victims; similarly, punishment by imprisonment is a net loss to society but is a negligible loss to victims, since they usually pay a negligible part of imprisonment costs. This is why b_j is often less than or equal to zero, at the same time that b, the coefficient of social loss, is greater than or equal to zero.

Since b_j and f_j are determined primarily by public policy on punishments, the main decision variable directly controlled by j is p_j. If he chooses a p_j that minimizes L_j, the optimality condition analogous to

[58] An increase in C_k—O_j and C held constant—presumably helps solve offenses against j, because more of those against k would be solved.

[59] The expected private loss, unlike the expected social loss, is apt to have considerable variance because of the small number of independent offenses committed against any single person. If j were not risk neutral, therefore, L would have to be modified to include a term that depended on the distribution of $b_j p_j f_j O_j$.

Copyright © 2001. All Rights Reserved.

equation (22) is

$$H_j' + C_j' + C_{jp_j}\frac{\partial p_j}{\partial O_j} = -b_j p_j f_j\left(1 - \frac{1}{\varepsilon_{jp_j}}\right).^{60} \tag{32}$$

The elasticity ε_{jp_j} measures the effect of a change in p_j on the number of offenses committed against j. If $b_j < 0$, and if the left-hand side of equation (32), the marginal cost of changing O_j, were greater than zero, then (32) implies that $\varepsilon_{jp_j} > 1$. Since offenders can substitute among victims, ε_{jp_j} is probably much larger than ε_p, the response of the total number of offenses to a change in the average probability, p. There is no inconsistency, therefore, between a requirement from the optimality condition given by (22) that $\varepsilon_p < 1$ and a requirement from (32) that $\varepsilon_{jp_j} > 1$.

VII. Some Applications

A. Optimal Benefits

Our analysis of crime is a generalization of the economist's analysis of external harm or diseconomies. Analytically, the generalization consists in introducing costs of apprehension and conviction, which make the probability of apprehension and conviction an important decision variable, and in treating punishment by imprisonment and other methods as well as by monetary payments. A crime is apparently not so different analytically from any other activity that produces external harm and when crimes are punishable by fines, the analytical differences virtually vanish.

Discussions of external economies or advantages are usually perfectly symmetrical to those of diseconomies, yet one searches in vain for analogues to the law of torts and criminality. Generally, compensation cannot be collected for the external advantages as opposed to harm caused, and no public officials comparable to policemen and district attorneys apprehend and "convict" benefactors rather than offenders. Of course, there is

[60] I have assumed that

$$\frac{\partial C}{\partial p_j} = \frac{\partial C_k}{\partial p_j} = 0,$$

in other words, that j is too "unimportant" to influence other expenditures. Although usually reasonable, this does suggest a modification to the optimality conditions given by eqs. (21) and (22). Since the effects of public expenditures depend on the level of private ones, and since the public is sufficiently "important" to influence private actions, eq. (22) has to be modified to

$$D' + C' + C_p\frac{\partial p}{\partial O} + \sum_{i=1}^{n}\frac{dC}{dC_i}\frac{dC_i}{dp}\frac{\partial p}{\partial O} = -bpf\left(1 + \frac{1}{\varepsilon_p}\right), \tag{22'}$$

and similarly for eq. (21). "The" probability p is, of course, a weighted average of the p_j. Eq. (22′) incorporates the presumption that an increase in public expenditures would be partially thwarted by an induced decrease in private ones.

Copyright © 2001. All Rights Reseved.

public interest in benefactors: medals, prizes, titles, and other privileges have been awarded to military heroes, government officials, scientists, scholars, artists, and businessmen by public and private bodies. Among the most famous are Nobel Prizes, Lenin Prizes, the Congressional Medal of Honor, knighthood, and patent rights. But these are piecemeal efforts that touch a tiny fraction of the population and lack the guidance of any body of law that codifies and analyzes different kinds of advantages.

Possibly the explanation for this lacuna is that criminal and tort law developed at the time when external harm was more common than advantages, or possibly the latter have been difficult to measure and thus considered too prone to favoritism. In any case, it is clear that the asymmetry in the law does not result from any analytical asymmetry, for a formal analysis of advantages, benefits, and benefactors can be developed that is quite symmetrical to the analysis of damages, offenses, and offenders. A function $A(B)$, for example, can give the net social advantages from B benefits in the same way that $D(O)$ gives the net damages from O offenses. Likewise, $K(B, p_1)$ can give the cost of apprehending and rewarding benefactors, where p_1 is the probability of so doing, with K' and $K_p > 0$; $B(p_1, a, v)$ can give the supply of benefits, where a is the award per benefit and v represents other determinants, with $\partial B/\partial p_1$ and $\partial B/\partial a > 0$; and b_1 can be the fraction of a that is a net loss to society. Instead of a loss function showing the decrease in social income from offenses, there can be a profit function showing the increase in income from benefits:

$$\Pi = A(B) - K(B, p_1) - b_1 p_1 a B. \tag{33}$$

If Π is maximized by choosing appropriate values of p_1 and a, the optimality conditions analogous to equations (21) and (22) are

$$A' - K' = b_1 p_1 a \left(1 + \frac{1}{e_a}\right) \tag{34}$$

and

$$A' - K' - K_p \frac{\partial p_1}{\partial B} = b_1 p_1 a \left(1 + \frac{1}{e_p}\right), \tag{35}$$

where

$$e_a = \frac{\partial B}{\partial a} \frac{a}{B}$$

and

$$e_p = \frac{\partial B}{\partial p_1} \frac{p_1}{B}$$

are both greater than zero. The implications of these equations are related to and yet differ in some important respects from those discussed earlier for (21) and (22).

For example, if $b_1 > 0$, which means that a is not a pure transfer but costs society resources, clearly (34) and (35) imply that $e_p > e_a$, since both $K_p > 0$ and $\partial p_1/\partial B > 0$. This is analogous to the implication of (21) and

Copyright © 2001. All Rights Reserved.

(22) that $\varepsilon_p > \varepsilon_f$, but, while the latter implies that, at the margin, offenders are risk *preferrers*, the former implies that, at the margin, benefactors are risk *avoiders*.[61] Thus, while the optimal values of p and f would be in a region where "crime does not pay"—in the sense that the marginal income of criminals would be less than that available to them in less risky legal activities—the optimal values of p_1 and a would be where "benefits do pay"—in the same sense that the marginal income of benefactors would exceed that available to them in less risky activities. In this sense it "pays" to do "good" and does not "pay" to do "bad."

As an illustration of the analysis, consider the problem of rewarding inventors for their inventions. The function $A(B)$ gives the total social value of B inventions, and A' gives the marginal value of an additional one. The function $K(B, p_1)$ gives the cost of finding and rewarding inventors; if a patent system is used, it measures the cost of a patent office, of preparing applications, and of the lawyers, judges, and others involved in patent litigation.[62] The elasticities e_p and e_a measure the response of inventors to changes in the probability and magnitude of awards, while b_1 measures the social cost of the method used to award inventors. With a patent system, the cost consists in a less extensive use of an invention than would otherwise occur, and in any monopoly power so created.

Equations (34) and (35) imply that with any system having $b_1 > 0$, the smaller the elasticities of response of inventors, the smaller should be the probability and magnitude of awards. (The value of a patent can be changed, for example, by changing its life.) This shows the relevance of the controversy between those who maintain that most inventions stem from a basic desire "to know" and those who maintain that most stem from the prospects of financial awards, especially today with the emphasis on systematic investment in research and development. The former quite consistently usually advocate a weak patent system, while the latter equally consistently advocate its strengthening.

[61] The relation $e_p > e_a$ holds if, and only if,

$$\frac{\partial EU}{\partial p_1}\frac{p_1}{U} > \frac{\partial EU}{\partial a}\frac{a}{U}, \tag{1'}$$

where

$$EU = p_1 U(Y + a) + (1 - p_1)U(Y) \tag{2'}$$

(see the discussion on pp. 177–78). By differentiating eq. (2′), one can write (1′) as

$$p_1[U(Y + a) - U(Y)] > p_1 a U'(Y + a), \tag{3'}$$

or

$$\frac{U(Y + a) - U(Y)}{a} > U'(Y + a). \tag{4'}$$

But (4′) holds if everywhere $U'' < 0$ and does not hold if everywhere $U'' \geq 0$, which was to be proved.

[62] These costs are not entirely trivial: for example, in 1966 the U.S. Patent Office alone spent $34 million (see Bureau of the Budget, 1967), and much more was probably spent in preparing applications and in litigation.

Copyright © 2001. All Rights Reseved.

Even if A', the marginal value of an invention, were "sizeable," the optimal decision would be to abolish property rights in an invention, that is, to set $p_1 = 0$, if b_1 and K[63] were sufficiently large and/or the elasticities e_p and e_a sufficiently small. Indeed, practically all arguments to eliminate or greatly alter the patent system have been based either on its alleged costliness, large K or b_1, or lack of effectiveness, low e_p or e_a (see, for example, Plant, 1934, or Arrow, 1962).

If a patent system were replaced by a system of cash prizes, the elasticities of response would become irrelevant for the determination of optimal policies, because b_1 would then be approximately zero.[64] A system of prizes would, moreover, have many of the same other advantages that fines have in punishing offenders (see the discussion in Sec. V). One significant advantage of a patent system, however, is that it automatically "meters" A', that is, provides an award that is automatically positively related to A', while a system of prizes (or of fines and imprisonment) has to estimate A' (or D') independently and often somewhat arbitrarily.

B. The Effectiveness of Public Policy

The anticipation of conviction and punishment reduces the loss from offenses and thus increases social welfare by discouraging some offenders. What determines the increase in welfare, that is "effectiveness," of public efforts to discourage offenses? The model developed in Section III can be used to answer this question if social welfare is measured by income and if "effectiveness" is defined as a ratio of the maximum feasible increase in income to the increase if all offenses causing net damages were abolished by fiat. The maximum feasible increase is achieved by choosing optimal values of the probability of apprehension and conviction, p, and the size of punishments, f (assuming that the coefficient of social loss from punishment, b, is given).[65]

[63] Presumably one reason patents are not permitted on basic research is the difficulty (that is, cost) of discovering the ownership of new concepts and theorems.

[64] The right side of both (34) and (35) would vanish, and the optimality conditions would be

$$A' - K' = 0 \tag{34'}$$

and

$$A' - K' - K_p \frac{\partial p_1}{\partial B} = 0. \tag{35'}$$

Since these equations are not satisfied by any finite values of p_1 and a, there is a difficulty in allocating the incentives between p_1 and a (see the similar discussion for fines in Sec. V).

[65] In symbols, effectiveness is defined as

$$E = \frac{D(O_1) - [D(\hat{O}) + C(\hat{p}, \hat{O}) + b\hat{p}\hat{f}\hat{O}]}{D(O_1) - D(O_2)},$$

where $\hat{p}$, $\hat{f}$, and $\hat{O}$ are optimal values, O_1 offenses would occur if $p = f = 0$, and O_2 is the value of O that minimizes D.

Copyright © 2001. All Rights Reserved.

Effectiveness so defined can vary between zero and unity and depends essentially on two behavioral relations: the costs of apprehension and conviction and the elasticities of response of offenses to changes in p and f. The smaller these costs or the greater these elasticities, the smaller the cost of achieving any given reduction in offenses and thus the greater the effectiveness. The elasticities may well differ considerably among different kinds of offenses. For example, crimes of passion, like murder or rape, or crimes of youth, like auto theft, are often said to be less responsive to changes in p and f than are more calculating crimes by adults, like embezzlement, antitrust violation, or bank robbery. The elasticities estimated by Smigel (1965) and Ehrlich (1967) for seven major felonies do differ considerably but are not clearly smaller for murder, rape, auto theft, and assault than for robbery, burglary, and larceny.[66]

Probably effectiveness differs among offenses more because of differences in the costs of apprehension and conviction than in the elasticities of response. An important determinant of these costs, and one that varies greatly, is the time between commission and detection of an offense.[67] For the earlier an offense is detected, the earlier the police can be brought in and the more likely that the victim is able personally to identify the offender. This suggests that effectiveness is greater for robbery than for a related felony like burglary, or for minimum-wage and fair-employment legislation than for other white-collar legislation like antitrust and public-utility regulation.[68]

C. A Theory of Collusion

The theory developed in this essay can be applied to any effort to preclude certain kinds of behavior, regardless of whether the behavior is "unlawful." As an example, consider efforts by competing firms to collude in order to obtain monopoly profits. Economists lack a satisfactory theory of the determinants of price and output policies by firms in an industry, a theory that could predict under what conditions perfectly competitive, monopolistic, or various intermediate kinds of behavior would emerge. One by-product of our approach to crime and punishment is a theory of collusion that appears to fill a good part of this lacuna.[69]

[66] A theoretical argument that also casts doubt on the assertion that less "calculating" offenders are less responsive to changes in p and f can be found in Becker (1962).

[67] A study of crimes in parts of Los Angeles during January, 1966, found that "more than half the arrests were made within 8 hours of the crime, and almost two-thirds were made within the first week" (President's Commission 1967*e*, p. 8).

[68] Evidence relating to the effectiveness of actual, which are not necessarily optimal, penalties for these white-collar crimes can be found in Stigler (1962, 1966), Landes (1966), and Johnson (1967).

[69] Jacob Mincer first suggested this application to me.

Copyright © 2001. All Rights Reseved.

The gain to firms from colluding is positively related to the elasticity of their marginal cost curves and is inversely related to the elasticity of their collective demand curve. A firm that violates a collusive arrangement by pricing below or producing more than is specified can be said to commit an "offense" against the collusion. The resulting harm to the collusion would depend on the number of violations and on the elasticities of demand and marginal cost curves, since the gain from colluding depends on these elasticities.

If violations could be eliminated without cost, the optimal solution would obviously be to eliminate all of them and to engage in pure monopoly pricing. In general, however, as with other kinds of offenses, there are two costs of eliminating violations. There is first of all the cost of discovering violations and of "apprehending" violators. This cost is greater the greater the desired probability of detection and the greater the number of violations. Other things the same, the latter is usually positively related to the number of firms in an industry, which partly explains why economists typically relate monopoly power to concentration. The cost of achieving a given probability of detection also depends on the number of firms, on the number of customers, on the stability of customer buying patterns, and on government policies toward collusive arrangements (see Stigler, 1964).

Second, there is the cost to the collusion of punishing violators. The most favorable situation is one in which fines could be levied against violators and collected by the collusion. If fines and other legal recourse are ruled out, methods like predatory price-cutting or violence have to be used, and they hurt the collusion as well as violators.

Firms in a collusion are assumed to choose probabilities of detection, punishments to violators, and prices and outputs that minimize their loss from violations, which would at the same time maximize their gain from colluding. Optimal prices and outputs would be closer to the competitive position the more elastic demand curves were, the greater the number of sellers and buyers, the less transferable punishments were, and the more hostile to collusion governments were. Note that misallocation of resources could not be measured simply by the deviation of actual from competitive outputs but would depend also on the cost of enforcing collusions. Note further, and more importantly, that this theory, unlike most theories of pricing, provides for continuous variation, from purely competitive through intermediate situations to purely monopolistic pricing. These situations differ primarily because of differences in the "optimal" number of violations, which in turn are related to differences in the elasticities, concentrations, legislation, etc., already mentioned.

These ideas appear to be helpful in understanding the relative success of collusions in illegal industries themselves! Just as firms in legal industries have an incentive to collude to raise prices and profits, so too do

Copyright © 2001. All Rights Reserved.

firms producing illegal products, such as narcotics, gambling, prostitution, and abortion. The "syndicate" is an example of a presumably highly successful collusion that covers several illegal products.[70] In a country like the United States that prohibits collusions, those in illegal industries would seem to have an advantage, because force and other illegal methods could be used against violators without the latter having much legal recourse. On the other hand, in countries like prewar Germany that legalized collusions, those in legal industries would have an advantage, because violators could often be legally prosecuted. One would predict, therefore, from this consideration alone, relatively more successful collusions in illegal industries in the United States, and in legal ones in prewar Germany.

VIII. Summary and Concluding Remarks

This essay uses economic analysis to develop optimal public and private policies to combat illegal behavior. The public's decision variables are its expenditures on police, courts, etc., which help determine the probability (p) that an offense is discovered and the offender apprehended and convicted, the size of the punishment for those convicted (f), and the form of the punishment: imprisonment, probation, fine, etc. Optimal values of these variables can be chosen subject to, among other things, the constraints imposed by three behavioral relations. One shows the damages caused by a given number of illegal actions, called offenses (O), another the cost of achieving a given p, and the third the effect of changes in p and f on O.

"Optimal" decisions are interpreted to mean decisions that minimize the social loss in income from offenses. This loss is the sum of damages, costs of apprehension and conviction, and costs of carrying out the punishments imposed, and can be minimized simultaneously with respect to p, f, and the form of f unless one or more of these variables is constrained by "outside" considerations. The optimality conditions derived from the minimization have numerous interesting implications that can be illustrated by a few examples.

If carrying out the punishment were costly, as it is with probation, imprisonment, or parole, the elasticity of response of offenses with respect to a change in p would generally, in equilibrium, have to exceed its response to a change in f. This implies, if entry into illegal activities can be explained by the same model of choice that economists use to explain entry into legal activities, that offenders are (at the margin) "risk preferrers." Consequently, illegal activities "would not pay" (at the margin)

[70] An interpretation of the syndicate along these lines is also found in Schilling (1967).

Copyright © 2001. All Rights Reseved.

in the sense that the real income received would be less than what could be received in less risky legal activities. The conclusion that "crime would not pay" is an optimality condition and not an implication about the efficiency of the police or courts; indeed, it holds for any level of efficiency, as long as optimal values of p and f appropriate to each level are chosen.

If costs were the same, the optimal values of both p and f would be greater, the greater the damage caused by an offense. Therefore, offenses like murder and rape should be solved more frequently and punished more severely than milder offenses like auto theft and petty larceny. Evidence on actual probabilities and punishments in the United States is strongly consistent with this implication of the optimality analysis.

Fines have several advantages over other punishments: for example, they conserve resources, compensate society as well as punish offenders, and simplify the determination of optimal p's and f's. Not surprisingly, fines are the most common punishment and have grown in importance over time. Offenders who cannot pay fines have to be punished in other ways, but the optimality analysis implies that the monetary value to them of these punishments should generally be less than the fines.

Vengeance, deterrence, safety, rehabilitation, and compensation are perhaps the most important of the many desiderata proposed throughout history. Next to these, minimizing the social loss in income may seem narrow, bland, and even quaint. Unquestionably, the income criterion can be usefully generalized in several directions, and a few have already been suggested in the essay. Yet one should not lose sight of the fact that it is more general and powerful than it may seem and actually includes more dramatic desiderata as special cases. For example, if punishment were by an optimal fine, minimizing the loss in income would be equivalent to compensating "victims" fully and would eliminate the "alarm" that so worried Bentham; or it would be equivalent to deterring all offenses causing great damage if the cost of apprehending, convicting, and punishing these offenders were relatively small. Since the same could also be demonstrated for vengeance or rehabilitation, the moral should be clear: minimizing the loss in income is actually very general and thus is *more useful* than these catchy and dramatic but inflexible desiderata.

This essay concentrates almost entirely on determining optimal policies to combat illegal behavior and pays little attention to actual policies. The small amount of evidence on actual policies that I have examined certainly suggests a positive correspondence with optimal policies. For example, it is found for seven major felonies in the United States that more damaging ones are penalized more severely, that the elasticity of response of offenses to changes in p exceeds the response to f, and that both are usually less than unity, all as predicted by the optimality analysis. There are, however, some discrepancies too: for example, the actual tradeoff between imprisonment and fines in different statutes is frequently less, rather than the

Copyright © 2001. All Rights Reserved.

predicted more, favorable to those imprisoned. Although many more studies of actual policies are needed, they are seriously hampered on the empirical side by grave limitations in the quantity and quality of data on offenses, convictions, costs, etc., and on the analytical side by the absence of a reliable theory of political decision-making.

Reasonable men will often differ on the amount of damages or benefits caused by different activities. To some, any wage rates set by competitive labor markets are permissible, while to others, rates below a certain minimum are violations of basic rights; to some, gambling, prostitution, and even abortion should be freely available to anyone willing to pay the market price, while to others, gambling is sinful and abortion is murder. These differences are basic to the development and implementation of public policy but have been excluded from my inquiry. I assume consensus on damages and benefits and simply try to work out rules for an optimal implementation of this consensus.

The main contribution of this essay, as I see it, is to demonstrate that optimal policies to combat illegal behavior are part of an optimal allocation of resources. Since economics has been developed to handle resource allocation, an "economic" framework becomes applicable to, and helps enrich, the analysis of illegal behavior. At the same time, certain unique aspects of the latter enrich economic analysis: some punishments, such as imprisonments, are necessarily non-monetary and are a cost to society as well as to offenders; the degree of uncertainty is a decision variable that enters both the revenue and cost functions; etc.

Lest the reader be repelled by the apparent novelty of an "economic" framework for illegal behavior, let him recall that two important contributors to criminology during the eighteenth and nineteenth centuries, Beccaria and Bentham, explicitly applied an economic calculus. Unfortunately, such an approach has lost favor during the last hundred years, and my efforts can be viewed as a resurrection, modernization, and thereby I hope improvement on these much earlier pioneering studies.

Mathematical Appendix

This Appendix derives the effects of changes in various parameters on the optimal values of p and f. It is assumed throughout that $b > 0$ and that equilibrium occurs where

$$\frac{\partial D}{\partial O} + \frac{\partial C}{\partial O} + \frac{\partial C}{\partial p}\frac{\partial p}{\partial O} = D' + C' + C_p \frac{\partial p}{\partial O} > 0;$$

the analysis could easily be extended to cover negative values of b and of this marginal cost term. The conclusion in the text (Sec. II) that $D'' + C'' > 0$ is relied on here. I take it to be a reasonable first approximation that the elasticities of O with respect to p or f are constant. At several places a sufficient

Copyright © 2001. All Rights Reseved.

condition for the conclusions reached is that

$$C_{pO} = C_{Op} = \frac{\partial^2 C}{\partial p \partial O} = \frac{\partial^2 C}{\partial O \partial p}$$

is "small" relative to some other terms. This condition is utilized in the form of a strong assumption that $C_{pO} = 0$, although I cannot claim any supporting intuitive or other evidence.

The social loss in income from offenses has been defined as

$$L = D(O) + C(O, p) + bpfO. \tag{A1}$$

If b and p were fixed, the value of f that minimized L would be found from the necessary condition

$$\frac{\partial L}{\partial f} = 0 = (D' + C')\frac{\partial O}{\partial f} + bpf(1 - E_f)\frac{\partial O}{\partial f}, \tag{A2}$$

or

$$0 = D' + C' + bpf(1 - E_f), \tag{A3}$$

if

$$\frac{\partial O}{\partial f} = O_f \neq 0,$$

where

$$E_f = \frac{-\partial f}{\partial O}\frac{O}{f}.$$

The sufficient condition would be that $\partial^2 L/\partial f^2 > 0$; using $\partial L/\partial f = 0$ and E_f is constant, this condition becomes

$$\frac{\partial^2 L}{\partial f^2} = (D'' + C'')O_f^2 + bp(1 - E_f)O_f > 0, \tag{A4}$$

or

$$\Delta \equiv D'' + C'' + bp(1 - E_f)\frac{1}{O_f} > 0. \tag{A5}$$

Since $D' + C' > 0$, and b is not less than zero, equation (A3) implies that $E_f > 1$. Therefore Δ would be greater than zero, since we are assuming that $D'' + C'' > 0$; and $\hat{f}$, the value of f satisfying (A3), would minimize (locally) the loss L.

Suppose that D' is positively related to an exogenous variable α. The effect of a change in α on $\hat{f}$ can be found by differentiating equation (A3):

$$D'_\alpha + (D'' + C'')O_f\frac{d\hat{f}}{d\alpha} + bp(1 - E_f)\frac{d\hat{f}}{d\alpha} = 0,$$

or

$$\frac{d\hat{f}}{d\alpha} = \frac{-D'_\alpha(1/O_f)}{\Delta}. \tag{A6}$$

Since $\Delta > 0$, $O_f < 0$, and by assumption $D'_\alpha > 0$, then

$$\frac{d\hat{f}}{d\alpha} = \frac{+}{+} > 0. \tag{A7}$$

In a similar way it can be shown that, if C' is positively related to an exogenous variable β,

$$\frac{d\hat{f}}{d\beta} = \frac{-C'_\beta(1/O_f)}{\Delta} = \frac{+}{+} > 0. \tag{A8}$$

Copyright © 2001. All Rights Reserved.

If b is positively related to γ, then

$$(D'' + C'')O_f \frac{d\hat{f}}{d\gamma} + bp(1 - E_f)\frac{d\hat{f}}{d\gamma} + pf(1 - E_f)b\gamma = 0,$$

or

$$\frac{d\hat{f}}{d\gamma} = \frac{-b_\gamma pf(1 - E_f)(1/O_f)}{\Delta}. \tag{A9}$$

Since $1 - E_f < 0$, and by assumption $b_\gamma > 0$,

$$\frac{d\hat{f}}{d\gamma} = \frac{-}{+} < 0. \tag{A10}$$

Note that since $1/E_f < 1$,

$$\frac{d(p\hat{f}O)}{d\gamma} < 0. \tag{A11}$$

If E_f is positively related to δ, then

$$\frac{d\hat{f}}{d\delta} = \frac{E_{f\delta}bpf(1/O_f)}{\Delta} = \frac{-}{+} < 0. \tag{A12}$$

Since the elasticity of O with respect to f equals

$$\epsilon_f = -O_f \frac{f}{O} = \frac{1}{E_f},$$

by (A12), a reduction in ϵ_f would reduce $\hat{f}$.

Suppose that p is related to the exogeneous variable r. Then the effect of a shift in r on $\hat{f}$ can be found from

$$(D'' + C'')O_f \frac{d\hat{f}}{dr} + (D'' + C'')O_p p_r + C_{pO}p_r + bp(1 - E_f)\frac{\partial \hat{f}}{\partial r} + bf(1 - E_f)p_r = 0,$$

or

$$\frac{d\hat{f}}{dr} = \frac{-(D'' + C'')O_p(1/O_f)p_r - bf(1 - E_f)p_r(1/O_f)}{\Delta}, \tag{A13}$$

since by assumption $C_{pO} = 0$. Since $O_p < 0$, and $(D'' + C'') > 0$,

$$\frac{d\hat{f}}{dr} = \frac{(-) + (-)}{+} = \frac{-}{+} < 0. \tag{A14}$$

If f rather than p were fixed, the value of p that minimizes L, $\hat{p}$, could be found from

$$\frac{\partial L}{\partial p} = \left[D' + C' + C_p \frac{1}{O_p} + bpf(1 - E_p)\right]O_p = 0, \tag{A15}$$

as long as

$$\frac{\partial^2 L}{\partial p^2} = \left[(D'' + C'')O_p + C'_p + C_{pp}\frac{1}{O_p} + C_{pO} + C_p \frac{\partial^2 p}{\partial O \partial p} + bf(1 - E_p)\right]O_p > 0. \tag{A16}$$

Copyright © 2001. All Rights Reseved.

Since $C'_p = C_{pO} = 0$, (A16) would hold if

$$\Delta' \equiv D'' + C'' + C_{pp}\frac{1}{O_p^2} + C_p\frac{1}{O_p}\frac{\partial^2 p}{\partial O \partial p} + bf(1 - E_p)\frac{1}{O_p} > 0. \tag{A17}$$

It is suggested in Section II that C_{pp} is generally greater than zero. If, as assumed,

$$D' + C' + C_p\frac{1}{O_p} > 0,$$

equation (A15) implies that $E_p > 1$ and thus that

$$bf(1 - E_p)\frac{1}{O_p} > 0.$$

If E_p were constant, $\partial^2 p/\partial O \partial p$ would be negative,[71] and, therefore, $C_p(1/O_p)(\partial^2 p/\partial O \partial p)$ would be positive. Hence, none of the terms of (A17) are negative, and a value of p satisfying equation (A15) would be a local minimum.

The effects of changes in different parameters on $\hat{p}$ are similar to those already derived for $\hat{f}$ and can be written without comment:

$$\frac{d\hat{p}}{d\alpha} = \frac{-D'_\alpha(1/O_p)}{\Delta'} > 0, \tag{A18}$$

$$\frac{d\hat{p}}{d\beta} = \frac{-C'_\beta(1/O_p)}{\Delta'} > 0, \tag{A19}$$

and

$$\frac{d\hat{p}}{d\gamma} = \frac{-b_\gamma pf(1 - E_p)(1/O_p)}{\Delta'} < 0. \tag{A20}$$

If E_p is positively related to δ',

$$\frac{d\hat{p}}{d\delta'} = \frac{E_{p\delta'}bpf(1/O_p)}{\Delta'} < 0. \tag{A21}$$

If C_p were positively related to the parameter s, the effect of a change in s on $\hat{p}$ would equal

$$\frac{d\hat{p}}{ds} = \frac{-C_{ps}(1/O_p^2)}{\Delta'} < 0. \tag{A22}$$

If f were related to the exogenous parameter t, the effect of a change in t on $\hat{p}$ would be given by

$$\frac{d\hat{p}}{dt} = \frac{-(D'' + C'')O_p f_t(1/O_p) - bf(1 - E_p)f_t(1/O_p) - C_p(\partial^2 p/\partial O \partial f)f_t(1/O_p)}{\Delta'} < 0 \tag{A23}$$

(with $C_{pO} = 0$), since all the terms in the numerator are negative.

If both p and f were subject to control, L would be minimized by choosing optimal values of both variables simultaneously. These would be given by the

[71] If E_p and E_f are constants, $O = kp^{-a}f^{-b}$, where $a = 1/E_p$ and $b = 1/E_f$. Then

$$\frac{\partial p}{\partial O} = -\frac{1}{ka}p^{a+1}f^b,$$

and

$$\frac{\partial^2 p}{\partial O \partial p} = \frac{-(a+1)}{ka}p^a f^b < 0.$$

Copyright © 2001 All Rights Reserved

solutions to the two first-order conditions, equations (A2) and (A15), assuming that certain more general second-order conditions were satisfied. The effects of changes in various parameters on these optimal values can be found by differentiating both first-order conditions and incorporating the restrictions of the second-order conditions.

The values of p and f satisfying (A2) and (A15), $\hat{p}$ and $\hat{f}$, minimize L if

$$L_{pp} > 0, \; L_{ff} > 0, \tag{A24}$$

and

$$L_{pp}L_{ff} > L_{fp}^2 = L_{pf}^2. \tag{A25}$$

But $L_{pp} = O_p^2\Delta'$, and $L_{ff} = O_f^2\Delta$, and since both Δ' and Δ have been shown to be greater than zero, (A24) is proved already, and only (A25) remains. By differentiating L_f with respect to p and utilizing the first-order condition that $L_f = 0$, one has

$$L_{fp} = O_f O_p[D'' + C'' + bf(1 - E_f)p_O] = O_f O_p \Sigma, \tag{A26}$$

where Σ equals the term in brackets. Clearly $\Sigma > 0$.

By substitution, (A25) becomes

$$\Delta\Delta' > \Sigma^2, \tag{A27}$$

and (A27) holds if Δ and Δ' are both greater than Σ. $\Delta > \Sigma$ means that

$$D'' + C'' + bp(1 - E_f)f_O > D'' + C'' + bf(1 - E_f)p_O, \tag{A28}$$

or

$$\frac{bfp}{O}(1 - E_f)E_f < \frac{bpf}{O}(1 - E_f)E_p. \tag{A29}$$

Since $1 - E_f < 0$, (A29) implies that

$$E_f > E_p, \tag{A30}$$

which necessarily holds given the assumption that $b > 0$; prove this by combining the two first-order conditions (A2) and (A15). $\Delta' > \Sigma$ means that

$$D'' + C'' + C_{pp}p_O^2 + C_p p_O p_{Op} + bf(1 - E_p)p_O > D'' + C'' + bf(1 - E_f)p_O. \tag{A31}$$

Since $C_{pp}p_O^2 > 0$, and $p_O < 0$, this necessarily holds if

$$C_p p p_{Op} + bpf(1 - E_p) < bpf(1 - E_f). \tag{A32}$$

By eliminating $D' + C'$ from the first-order conditions (A2) and (A15) and by combining terms, one has

$$C_p p_O - bpf(E_p - E_f) = 0. \tag{A33}$$

By combining (A32) and (A33), one gets the condition

$$C_p p p_{Op} < C_p p_O, \tag{A34}$$

or

$$E_{p_O,p} = \frac{p}{p_O}\frac{\partial p_O}{\partial p} > 1. \tag{A35}$$

It can be shown that

$$E_{p_O,p} = 1 + \frac{1}{E_p} > 1, \tag{A36}$$

and, therefore, (A35) is proven.

Copyright © 2001. All Rights Reseved.

It has now been proved that the values of p and f that satisfy the first-order conditions (A2) and (A15) do indeed minimize (locally) L. Changes in different parameters change these optimal values, and the direction and magnitude can be found from the two linear equations

$$O_f\Delta\frac{\partial\tilde{f}}{\partial z} + O_p\Sigma\frac{\partial\tilde{p}}{\partial z} = C_1$$

and (A37)

$$O_f\Sigma\frac{\partial\tilde{f}}{\partial z} + O_p\Delta'\frac{\partial\tilde{p}}{\partial z} = C_2.$$

By Cramer's rule,

$$\frac{\partial\tilde{f}}{\partial z} = \frac{C_1O_p\Delta' - C_2O_p\Sigma}{O_pO_f(\Delta\Delta' - \Sigma^2)} = \frac{O_p(C_1\Delta' - C_2\Sigma)}{+}, \tag{A38}$$

$$\frac{\partial\tilde{p}}{\partial z} = \frac{C_2O_f\Delta - C_1O_f\Sigma}{O_pO_f(\Delta\Delta' - \Sigma^2)} = \frac{O_f(C_2\Delta - C_1\Sigma)}{+}, \tag{A39}$$

and the signs of both derivatives are the same as the signs of the numerators.

Consider the effect of a change in D' resulting from a change in the parameter α. It is apparent that $C_1 = C_2 = -D'_\alpha$, and by substitution

$$\frac{\partial\tilde{f}}{\partial\alpha} = \frac{-O_pD'_\alpha(\Delta' - \Sigma)}{+} = \frac{+}{+} > 0 \tag{A40}$$

and

$$\frac{\partial\tilde{p}}{\partial\alpha} = \frac{-O_pD'_\alpha(\Delta - \Sigma)}{+} = \frac{+}{+} > 0, \tag{A41}$$

since O_f and $O_p < 0$, $D'_\alpha > 0$, and Δ and $\Delta' > \Sigma$.

Similarly, if C' is changed by a change in β, $C_1 = C_2 = -C'_\beta$,

$$\frac{\partial\tilde{f}}{\partial\beta} = \frac{-O_pC'_\beta(\Delta' - \Sigma)}{+} = \frac{+}{+} > 0, \tag{A42}$$

and

$$\frac{\partial\tilde{p}}{\partial\beta} = \frac{-O_fC'_\beta(\Delta - \Sigma)}{+} = \frac{+}{+} > 0. \tag{A43}$$

If E_f is changed by a change in δ, $C_1 = E_{f\delta}bpf$, $C_2 = 0$,

$$\frac{\partial\tilde{f}}{\partial\delta} = \frac{O_pE_fbpf\Delta'}{+} = \frac{-}{+} < 0, \tag{A44}$$

and

$$\frac{\partial\tilde{p}}{\partial\delta} = \frac{-O_fE_fbpf\Sigma}{+} = \frac{+}{+} > 0. \tag{A45}$$

Similarly, if E_p is changed by a change in δ', $C_1 = 0$, $C_2 = E_{p\delta'}bpf$,

$$\frac{\partial\tilde{f}}{\partial\delta'} = -\frac{O_pE_{p\delta'}bpf\Sigma}{+} = \frac{+}{+} > 0, \tag{A46}$$

and

$$\frac{\partial\tilde{p}}{\partial\delta'} = \frac{O_fE_{p\delta'}bpf\Delta}{+} = \frac{-}{+} < 0. \tag{A47}$$

Copyright © 2001. All Rights Reserved.

If b is changed by a change in γ, $C_1 = -b_\gamma pf(1 - E_f)$, $C_2 = -b_\gamma pf(1 - E_p)$, and

$$\frac{\partial \tilde{f}}{\partial \gamma} = \frac{-O_p b_\gamma pf[(1 - E_f)\Delta' - (1 - E_p)\Sigma]}{+} = \frac{-}{+} < 0, \tag{A48}$$

since $E_f > E_p > 1$ and $\Delta' > \Sigma$; also,

$$\frac{\partial \tilde{p}}{\partial \gamma} = \frac{-O_f b_\gamma pf[(1 - E_p)\Delta - (1 - E_f)\Sigma]}{+} = \frac{+}{+} > 0, \tag{A49}$$

for it can be shown that $(1 - E_p)\Delta > (1 - E_f)\Sigma$.[72] Note that when f is held constant the optimal value of p is decreased, not increased, by an increase in γ.

If C_p is changed by a change in s, $C_2 = -p_O C_{ps}$, $C_1 = 0$,

$$\frac{\partial \tilde{f}}{\partial s} = \frac{O_p p_O C_{ps}\Sigma}{+} = \frac{C_{ps}\Sigma}{+} = \frac{+}{+} > 0, \tag{A50}$$

and

$$\frac{\partial \tilde{p}}{\partial s} = \frac{-O_f p_O C_{ps}\Delta}{+} = \frac{-}{+} < 0. \tag{A51}$$

[72] The term $(1 - E_p)\Delta$ would be greater than $(1 - E_f)\Sigma$ if

$$(D'' + C'')(1 - E_p) + bp(1 - E_f)(1 - E_p)f_O > (D'' + C'')(1 - E_f) + bf(1 - E_f)^2 p_O,$$

or

$$(D'' + C'')(E_f - E_p) > -\frac{bpf}{O}(1 - E_f)\left[(1 - E_p)\frac{f_O O}{f} - (1 - E_f)\frac{p_O O}{p}\right],$$

$$(D'' + C'')(E_f - E_p) > -\frac{bpf}{O}(1 - E_f)[(1 - E_p)(E_f) - (1 - E_f)E_p],$$

$$(D'' + C'')(E_f - E_p) > -\frac{bpf}{O}(1 - E_f)(E_f - E_p).$$

Since the left-hand side is greater than zero, and the right-hand side is less than zero, the inequality must hold.

References

Arrow, Kenneth J. "Economic Welfare and Allocation of Resources for Invention," in National Bureau Committee for Economic Research. *The Rate and Direction of Inventive Activity: Economic and Social Factors.* Princeton, N.J.: Princeton Univ. Press (for the Nat. Bureau of Econ. Res.), 1962.

Becker, Gary S. "Irrational Behavior and Economic Theory," *J.P.E.*, Vol. LXX (February, 1962).

———. "A Theory of the Allocation of Time," *Econ. J.*, Vol. LXXV (September, 1965).

Bentham, Jeremy. *Theory of Legislation.* New York: Harcourt Brace Co., 1931.

Bureau of the Budget. *The Budget of United States Government, 1968, Appendix.* Washington: U.S. Government Printing Office, 1967.

Copyright © 2001. All Rights Reseved.

Bureau of Prisons. *Prisoners Released from State and Federal Institutions.* ("National Prisoner Statistics.") Washington: U.S. Dept. of Justice, 1960.

———. *Characteristics of State Prisoners, 1960.* ("National Prisoner Statistics.") U.S. Dept. of Justice, n.d.

———. *Federal Prisons, 1960.* Washington: U.S. Dept. of Justice, 1961.

Cagan, Phillip. *Determinants and Effects of Changes in the Stock of Money, 1875–1960.* New York: Columbia Univ. Press (for the Nat. Bureau of Econ. Res.), 1965.

"Criminal Safeguards and the Punitive Damages Defendant," *Univ. Chicago Law Rev.*, Vol. XXXIV (Winter, 1967).

Ehrlich, Isaac. "The Supply of Illegitimate Activities." Unpublished manuscript, Columbia Univ., New York, 1967.

Federal Bureau of Investigation. *Uniform Crime Reports for the United States.* Washington: U.S. Dept. of Justice, 1960.

———. *Ibid.*, 1961.

Harper, F. V., and James, F. *The Law of Torts*, Vol. II. Boston: Little-Brown & Co., 1956.

Johnson, Thomas. "The Effects of the Minimum Wage Law." Unpublished Ph.D. dissertation, Columbia Univ., New York, 1967.

Kleinman, E. "The Choice between Two 'Bads'—Some Economic Aspects of Criminal Sentencing." Unpublished manuscript, Hebrew Univ., Jerusalem, 1967.

Landes, William. "The Effect of State Fair Employment Legislation on the Economic Position of Nonwhite Males." Unpublished Ph.D. dissertation, Columbia Univ., New York, 1966.

Laws of New York, Vol. II (1965).

Marshall, Alfred. *Principles of Economics.* 8th ed. New York: Macmillan Co., 1961.

Plant, A. "The Economic Theory concerning Patents for Inventions," *Economica*, Vol. I (February, 1934).

Pollock, F., and Maitland, F. W. *The History of English Law.* Vol. II. 2d ed. Cambridge: Cambridge Univ. Press, 1952.

President's Commission on Law Enforcement and Administration of Justice. *The Challenge of Crime in a Free Society.* Washington: U.S. Government Printing Office, 1967(*a*).

———. *Corrections.* ("Task Force Reports.") Washington: U.S. Government Printing Office, 1967(*b*).

———. *The Courts.* ("Task Force Reports.") Washington: U.S. Government Printing Office, 1967(*c*).

———. *Crime and Its Impact—an Assessment.* ("Task Force Reports.") Washington: U.S. Government Printing Office, 1967(*d*).

———. *Science and Technology.* ("Task Force Reports.") Washington: U.S. Government Printing Office, 1967(*e*).

Radzinowicz, L. *A History of English Criminal Law and Its Administration from 1750.* Vol. I. London: Stevens & Sons, 1948.

Schilling, T. C. "Economic Analysis of Organized Crime," in President's Commission on Law Enforcement and Administration of Justice. *Organized Crime.* ("Task Force Reports.") Washington: U.S. Government Printing Office, 1967.

Shawness, Lord. "Crime *Does* Pay because We Do Not Back Up the Police," *New York Times Magazine*, June 13. 1965.

Smigel, Arleen. "Crime and Punishment: An Economic Analysis." Unpublished M.A. thesis, Columbia Univ., New York, 1965.

Copyright © 2001. All Rights Reserved.

Stigler, George J. "What Can Regulators Regulate? The Case of Electricity," *J. Law and Econ.*, Vol. V (October, 1962).
———. "A Theory of Oligopoly," *J.P.E.*, Vol. LXXII (February, 1964).
———. "The Economic Effects of the Antitrust Laws," *J. Law and Econ.*, Vol. IX (October, 1966).
Sutherland, E. H. *Principles of Criminology*, 6th ed. Philadelphia: J. B. Lippincott Co., 1960.

Copyright © 2001. All Rights Reseved.