Some Results on Incentive Contracts with Applications to Education and Employment, Health Insurance, and Law Enforcement

By MILTON HARRIS AND ARTUR RAVIV*

When decision-making authority is delegated from one agent to another, contractual arrangements are often used to allocate resources and outputs. Such situations may be analyzed using the theory of principal-agent relationships. This theory seeks to characterize optimal contracts and explain observed arrangements. Examples which fit the "agency paradigm" include employer-employee, insurer-insured, and owner-manager relations. In this paper, we report some results which significantly extend the theory of agency to situations characterized by a divergence of incentives between the two parties and asymmetric information with opportunities for acquiring information. In addition, we discuss several applications of this theory.

The theory of optimal contracts under conditions of uncertainty has received considerable attention. Kenneth Arrow and Robert Wilson (1968) were concerned with the optimal sharing of purely exogenous risk. Wilson (1969) and Stephen Ross considered situations in which the risk could be affected by the actions of the agents. They analyze contracts which induce similar attitudes toward risk on the part of the agents, thus allowing the possibility that decentralized decision making will be optimal. Conditions under which such arrangements are indeed Pareto optimal are also investigated. In their models, incentive problems arise purely as a consequence of diverse attitudes toward risk among the agents. A. Michael Spence and Richard Zeckhauser, in the context of insurance contracts, introduced the problem of a divergence in incentives due to the action of an agent together with differential information among agents. More recently, Joseph Stiglitz (1975a) analyzed incentive contracts between employers and employees. With regard to differential information, both Spence-Zeckhauser and Stiglitz assume one of two extreme cases: either the agent's action is known by everyone with certainty (in which case there is no differential information), or no information about the agent's action is available to anyone except the agent himself. A somewhat intermediate case was analyzed by Robert Townsend in which the exact information possessed by one agent can be made available to the other at some cost.

Two important aspects of agency relationships are not fully explored in the literature on the theory of contracts. First, most agency relationships must deal with incentive problems which arise because the agent would prefer to work less, other things equal, while the principal is indifferent to the level of the agent's effort, other things (i.e., his share of the payoff) equal. This type of incentive problem is somewhat different from the one considered by Wilson (1969) and Ross in which a divergence in incentives results only from different attitudes toward risk. Second, in most instances an agent may acquire information about other agents' actions. This possibility was discussed in interesting papers by Armen Alchian and Harold Demsetz and by C. Michael Jensen and William Meckling. The quality of the information obtained through monitoring (or supervising) depends on the resources committed to this activity as well as on the available monitoring technology. Furthermore, as Stiglitz

*Carnegie-Mellon University. We would like to acknowledge helpful discussions with Ed Prescott and Rob Townsend, as well as suggestions of Stephen Ross, Martin Hellwig, an anonymous referee, and the managing editor of this *Review*.

points out, "... the amount (or quality) of supervision will affect both the optimal incentive scheme which will be used and the level of expected utility which the individual will attain" (1975a, p. 572). Consequently, the optimal incentive contract will depend on the available monitoring technology. In this paper, we explore these aspects of the agency problem.

Our analysis is based on a model in which there are two individuals: one, denoted the agent, takes an action which together with the realization of an exogenous random variable results in the payoff to be divided between the agent and the other individual, denoted the principal. Incentive problems arise because the agent has a disutility for the action while the principal does not. We distinguish two versions of this model. In the first, the agent is assumed to take his action without any information regarding the realization of the exogenous random state. In the second version, the agent is assumed to know the value of the random state before taking his action. In both versions, the object of the analysis is to discover the form of the Pareto optimal contract, that is, how the optimal sharing arrangement for the payoff depends on the observed variables. In particular, our analysis deals with the following issues: When would we expect to observe performance-contingent contracts, and what would be the form of such contracts; when performance is not observable, under what conditions would we expect contracts to depend on imperfect estimates of performance, what types of estimators would be used, and how would they be incorporated into the contract.

Our results are discussed in terms of the employer-employee relationship. They are, however, more general, and we discuss their implications for three other agency relationships.

The first application is to the analysis of employment contracts based on training or education and ability. Here we show under what conditions and in what form it is optimal to make employment contracts contingent on training or ability. In particular we show when the type of contract used in the "signaling literature" (see Spence, 1973, 1974, 1976; John Riley; Stiglitz, 1975b) is Pareto optimal.

The second application is to health insurance contracts. We show when indemnity insurance, that is, contracts in which the insurance payment depends only on the degree of illness, is optimal. We also show when the optimal policy provides payment contingent on the level of medical care chosen. The case in which the level of medical care chosen is not directly observable is also considered.

The third application of the analysis is to the problem of procuring the optimal amount of law enforcement, an issue which was addressed by Gary Becker and George Stigler. We exhibit conditions on the information structure under which the standard type of compensation arrangement for the enforcer (i.e., a salary) would lead to an inefficiently large degree of malfeasance or nonfeasance (shirking). We also show under what conditions these inefficiencies could be resolved by the use of contracts which we exhibit. In particular, we provide conditions under which some suggestions of Becker and Stigler would be optimal.

I. The Agency Model

In this section we describe the model with which we address the issues mentioned above (a more detailed and formal treatment may be found in the authors' working paper). In this model there are two individuals, "the principal" and "the agent." For concreteness, in describing the model and results, we refer to these individuals as the employer and worker, respectively.

The worker chooses a level of effort (or action) a, which together with the realization of some exogenous random variable θ determines the value of the worker's product x (the payoff). The random variable θ may be interpreted as the result of any exogenous uncertain event which affects the worker's productivity, for example, the weather, equipment failure, the price of the product, etc. We represent the relationship

among the value of the worker's output, the worker's effort, and the realization of the random variable by a production function X, that is,

$$x = X(a, \theta) \tag{1}$$

We assume that greater effort by the worker results in greater output for any value of θ (i.e., $X_1 > 0$ for all a, θ).

Two important cases may be distinguished regarding the information available to the worker when he decides upon his level of effort. First, in Model 1, we assume that he does *not* know the value of θ when he chooses his effort. For example, a sharecropper may not know what the weather will be when he plants his crop; a lawyer, when he prepares his case, may not know which judge will preside. Second, in Model 2, we assume that the worker *does* observe the value of θ and chooses his effort contingent on this observation. For example, a salesman may observe the state of demand before deciding how many calls to make.

The division of the product will be determined by a function which may depend on any variable which is observable by both parties, that is, whose value is known by both parties when the product is divided. This assumption rules out contracts which are not incentive compatible. The function and its list of arguments denoted $(S; z)$ will be called a *contract*. The value of $S(z)$ is interpreted as the worker's share of the product while the employer's share is $x - S(z)$. The product x is assumed always to be observable by both parties. In addition to x, the list z might include the effort a, provided, of course, that it is observable. If a is not observable, z might include an estimator, denoted α of a. For example, an estimator of the sharecropper's effort might be the amount of time he spends in the field. An estimator of the salesman's effort might be the number of miles he drives. When available, the estimator will be called a *monitor*, that is, a monitor is a random variable whose distribution is conditional on a. The class of available monitors is referred to as the *monitoring technology*.

Associated with the worker is a utility function U, whose first argument is the worker's share of the product and whose second argument is his effort. The worker is assumed to prefer less effort to more effort, other things equal. We therefore assume that

$$U_1 > 0; \qquad U_2 < 0 \tag{2}$$

We will often assume that the worker is risk averse, that is, that U is strictly concave in the first argument.

Given a contract $(S; z)$, in Model 1 the worker determines his effort by solving the following maximization problem:

$$\max_a E_\theta U[S(z), a] \tag{3}$$

In Model 2, the worker chooses a as a function of θ to solve

$$\max_a U[S(z), a] \tag{3'}$$

In both models, the effort chosen will depend on the functional form of S; in Model 2, it will also depend on the realization of the exogenous random variable.

Associated with the employer is a utility function V, which is a function *only* of his share of the product. We assume V is monotone increasing, concave (i.e., the employer is either risk neutral or risk averse). Given a contract $(S; z)$, and a level of effort a chosen by the worker according to either (3) or (3'), the employer's utility is

$$E_\theta V[X(a, \theta) - S(z)] \tag{4}$$

This concludes our presentation of the model. We turn in the next section to a description of our results.

II. Results

Our results characterize the Pareto optimal contracts under various assumptions concerning the availability of information. In particular, optimal contracts are investigated under alternative assumptions on the observability of the worker's effort and the random variable and on the existing tech-

nology for monitoring the worker's effort. We are interested in Pareto optimal contracts on the supposition that observed contracts will have the property that, given the availability of information, neither agent's expected utility can be increased without decreasing the expected utility of the other agent. Thus we seek to characterize the contracts which we expect to arise under various information structures. Our results are simply discussed here without formal derivations; these may be found in our earlier paper.

To begin, suppose that the realization of the random variable (also referred to as "the state of nature" or simply "the state") is freely observable by both employer and worker when the product is distributed. In this case, our first result implies that making the contract contingent on the worker's effort provides no gains over contracts which depend only on output and the observed value of θ. Therefore, if the state is freely observable, we would *not* expect to observe contracts contingent on worker effort. Thus *ex post* uncertainty as to the relationship between the effort and the product is essential if contracts based on worker performance are to be observed. For future reference we state this result as

PROPOSITION 1: *The expected utilities achieved under any contract which depends on the product, the effort, and the state can also be achieved under a contract which depends on the product and the state, but not on the effort. This result holds for both Models* 1 *and* 2.

The above result does not yield information as to the form of the optimal contract when the state is freely observable. This information is provided by

PROPOSITION 2: *The Pareto optimal contract which depends on the product and the state specifies a "standard" output contingent on the state. The worker receives an amount which depends on the standard output and perhaps on the state, plus the difference between the actual output and the standard. The employer receives an amount which depends only on the state and is therefore unaffected by the worker's level of effort. This result holds for both Models* 1 *and* 2.

To illustrate this result, consider the sharecropper example mentioned above. In this case, Propositions 1 and 2 imply that if the weather is the only exogenous random factor affecting the crop, we would expect to observe an arrangement in which the tenant "pays" the landlord an amount of output contingent only on the weather and keeps the remainder of the crop.

Note that Propositions 1 and 2 hold regardless of attitudes toward risk of the employer (landlord) and worker (tenant). Thus the determination of the variables on which the contract will depend (if all are observable) is independent of attitudes toward risk. For example, if output, effort, and the weather are all observable, the contract will depend only on output and weather. This result does depend, however, on the ability of both parties to agree *ex ante* on which types of weather are possible, what the probability attached to each type is, and how weather affects output. It also depends on the ability of both parties to observe *ex post* which type of weather occurred. Attitudes toward risk do play a role in determining the sharing function. For example, if the landlord is risk neutral, the contract will be such that if the tenant puts in sufficient effort to produce the (weather-contingent) standard output, then the tenant's share will be independent of the weather. If the tenant does not render the effort implicit in the standard output, however, his share may be weather-dependent. The landlord's share (i.e., rent) *will* depend on the weather, in general, if the tenant is risk averse.

Attitudes toward risk also play an important role in determining the variables on which the optimal contract will depend when the state is not observable, *ex post*. The next result states that risk aversion on the part of the worker is a necessary and

sufficient condition for contracts which depend explicitly on effort to be superior to contracts which depend only on the output. Intuitively, if the worker is risk neutral, it will be optimal for him to bear all the risk associated with uncertain productivity. In this case, all effects of the worker's performance are internalized, and thus incentive problems are resolved without the use of performance-contingent contracts. When the worker is risk averse, optimality requires some sharing of risk, and therefore the employer bears some of the consequences of the worker's choice of effort. In this case, performance-contingent contracts are superior to contracts based only on output.

PROPOSITION 3: (i) *If the worker is risk neutral, any contract which depends only on output and effort can be dominated (in the Pareto sense) by a contract which depends only on output.*

(ii) *If the worker is risk averse, any contract which depends only on output can be strictly dominated by a contract which depends on both output and effort.*

(iii) *If the worker is risk averse, any contract which depends only on output can be strictly dominated by a contract which depends only on output and the state and is of the form given in Proposition* 2.

These results hold for Models 1 and 2.[1] Next we characterize the form of Pareto optimal contracts which depend on output and effort.

PROPOSITION 4: *Any Pareto optimal contract in the class of contracts depending only on output and effort has the property that the worker's share depends only on the output, provided his effort meets a prespecified criterion. If not, he receives nothing. This result holds for Model* 1.

Thus when the worker's effort is freely observable, and the worker is risk averse, we would expect to observe contracts which stipulate a particular choice of effort by the worker. Since the worker will always choose to meet the requirement, we refer to this contract as a forcing contract.[2]

With respect to the sharecropping paradigm, these results imply that if the tenant is risk averse, and output and effort, *but not the weather,* are observable *ex post*,[3] we would expect the sharing arrangement to include a stipulation of how much effort the tenant is expected to expend. If the tenant puts in the required effort (and he will if his effort is perfectly observable), he will receive a share of the product. This share will depend on the product and the attitudes toward risk of the landlord and tenant. For example if the landlord is risk neutral (and not the tenant), the tenant will get a fixed wage independent of output (or the weather). If, on the other hand, the tenant (and not the landlord) is risk neutral, contracts stipulating the tenant's effort are unnecessary; we expect to observe the landlord receiving a fixed rent with the tenant keeping the residual.

We now turn to some results regarding the use of imperfect estimators of effort as a contingency affecting the distribution of the product. Clearly we would not expect to observe the use of imperfect monitors of effort when there are no gains to using actual effort. We have exhibited above two situations in which there are no gains to acquiring information about the worker's effort (even if such information is available costlessly). Moreover we have shown that in all other cases there are gains to acquiring information, that is, if information of "sufficient quality" can be obtained at a "sufficiently low price," then both individuals can be made better off. The two conditions

[1]For Model 2 the utility function of the worker is assumed to be separable and the production function is one-to-one in the state.

[2]This terminology was suggested to us by Ross.

[3]With respect to the sharecropping example, one might question the assumption that the tenant's effort is observable while the weather is not. We agree that these assumptions are not particularly appropriate to sharecropping. Our model is, however, applicable to a large class of situations, and for many of these, the assumptions are appropriate (see Section III). We use the sharecropping example here only for the purpose of illustrating all our results with a consistent example.

under which there are no gains to monitoring the worker's effort are (a) when the realization of θ, the exogenous random variable affecting output, is observable, and (b) when the worker is risk neutral. From a methodological point of view, these results imply that it is not possible to simplify the analysis of monitoring by assuming away either exogenous uncertainty or risk aversion on the part of the worker.

In general, even when there are gains to perfect information on effort, these gains may be impossible to realize through imperfect monitoring. The introduction of imperfect information on the worker's effort into a contract produces two opposing effects on the welfare of the parties to the contract. First, since the information is imperfect, additional uncertainty is introduced. Since both the employer and worker are risk averse, this additional uncertainty tends to reduce welfare. Second, inclusion of monitoring can motivate the worker to choose a level of effort which, neglecting the first effect, would make both parties better off.

Because the minimum (necessary) conditions for monitoring to be valuable appear to be very difficult to formulate, we explore several sets of sufficient conditions. These results are derived using forcing type contracts. This type of contract specifies that the worker is paid only if the monitoring reveals his effort to be "acceptable," and the size of the payment does not depend on the results of monitoring in any other way. The forcing contract thus requires that a decision be made based on the realization of the monitor. This statistical decision problem is the test of the following hypothesis: the worker's level of effort is one of a set of acceptable levels. The precision of the monitor, in this case, is summarized by the probabilities of type I and type II errors. The conditions on the monitoring technology referred to below are assumptions as to the availability of monitors with various probabilities of type I and type II errors.

Our results regarding the use of contracts based on imperfect monitoring of worker effort can be summarized as follows:

i) First, we establish general sufficient conditions under which the potential gains to monitoring may be realized.

ii) Under additional assumptions on the monitoring technology, we show that any Pareto optimal contract can be approximated to any degree of precision by a forcing type of contract.

iii) Under another set of assumptions on the monitoring technology and the worker's utility function, we show that all the results achievable under perfect information can be obtained even when information is imperfect, using a forcing contract.

The above results are obtained only for Model 1.

III. Applications

In this section, three applications of the results of the previous section are analyzed. These results are used to describe the Pareto optimal contractual arrangements, and therefore the contracts we would expect to observe in some interesting situations. First, we analyze employment contracts based on education, training, or ability. We compare our results to those of the signaling literature (for example, Spence, 1973, 1974). Second, we consider optimal health insurance contracts and explain indemnity insurance. Finally, the problem of optimal compensation of law enforcers is discussed. Here we relate our results to those of Becker and Stigler.

A. *Ability, Training, and Employment Contracts*

There has recently been considerable interest in the relationship between education and other "signals" and the allocation of labor in job markets. For example, Spence (1973, 1974, 1976) stresses the use of education as a signal for native ability in the hiring of employees (see also Riley, Stiglitz, 1975b). In this literature, a particular payment structure for the worker, based on the signal, is assumed. This structure involves paying the worker a fixed wage which de-

pends only on his education level. These papers provide no analysis to justify the use of such a payment schedule, nor do they consider the possibility that the use of the signal in this way may be Pareto inferior to some other contractual arrangement. The present section is devoted to a clarification of this issue. We characterize the Pareto optimal contract in several situations similar to those analyzed in the signaling literature. In particular we exhibit conditions under which an *optimal* contract will depend on the education level of the worker (as well as things like recommendations, transcripts, previous experience, etc.), even when education is not a perfect measure of acquired ability or productivity. Thus our results may be viewed as complementary with the signaling literature.

To apply our model to the present problem, we reinterpret the effort as the ability the worker has acquired to perform the task, and the state as his native ability. We assume that neither the worker nor his prospective employer knows the worker's native ability at the time the contract is agreed on. Finally, each worker is assumed to know his own *acquired* ability.

With this interpretation, our results of Section II are applicable and imply:

i) If native ability is known or observable *ex post,* it follows from Propositions 1 and 2 that the optimal contract specifies that the employer receives an amount which depends only on the worker's native ability. In particular, the employer specifies a standard output to be produced by the worker contingent on his native ability. The employer's payoff is a certain share (which may depend on the worker's native ability) of this "standard output" while the worker receives the remainder of the standard revenue plus any output generated in excess of the standard (or minus any shortfall of actual output from standard output). This arrangement is equivalent to specifying a standard level of acquired ability required from a worker in this particular job and allowing workers with different levels of acquired ability to participate and accept the full consequences if their acquired ability is different from the one required.

ii) If workers are risk neutral, from Proposition 3, the Pareto optimal contract will specify a given payoff to employers independent of output, the worker's acquired ability, or his native ability level. Therefore, the worker receives the entire output minus some constant, that is, the worker purchases the right to use the production function for a given price.

iii) If workers are risk averse, contracts which depend only on the output are inefficient relative to contracts which depend on the native ability and/or acquired ability of the worker as well as the output. Pareto-superior results can be obtained if the acquired ability is observable and is included in the contract. From Proposition 4 it follows that Pareto optimal contracts when only output and acquired ability are observable are of the form "workers with acquired ability level of (at least) a^* will be paid $S(x)$, others need not apply," where $S(x)$ is some function of the output. For example, if years of education is a perfect correlate of acquired ability, this result implies that it is Pareto inferior not to make employment and/or salaries contingent on the education of the worker. Even if acquired ability is not observable (nor is something perfectly correlated with it) there may be imperfect measures which are sufficiently precise to make using them Pareto efficient. Examples of such monitors include years of education, interviews, transcripts, recommendations, etc. In this case we have shown sufficient conditions under which it is Pareto inferior not to include such monitors of acquired ability. The contract used is of the forcing type and specifies a minimum required level of the observable monitor of acquired ability. The above analysis provides an explanation for compensation schedules which are functions of education level. According to our analysis it is optimal for workers to receive *fixed wages* contingent on education level when employers are risk neutral. If employers are risk averse, a worker's compensation will

depend on his output if he meets the education requirement.

B. *Health Insurance Contracts*

It has for some time been recognized that when insurance payments depend on a decision of the insured as well as the state of nature, then an optimal allocation of resources and risk will not be achieved by a simple arrangement in which the insured pays a given price (premium) in return for various payments contingent on the state of nature. This problem arises because the insured has an incentive to "overspend" on insured expenses. It has been called "moral hazard" in the insurance literature (see, for example, Arrow, chs. 5, 8, 9; Pauly, 1968). One way to mitigate the inefficiency is to impose some of the cost of medical care on the insured. Zeckhauser illustrates the tradeoff between risk spreading and incentives. He suggests that insurance payments should depend on the degree of illness as well as the cost of the associated health care. Pauly (1971) refers to this type of insurance as "indemnity" insurance. The results of Section II can be applied to the case of health insurance to show under what conditions moral hazard causes inefficiencies and how appropriate contractual arrangements such as indemnity insurance can resolve this problem. Therefore, the results of Zeckhauser and Pauly (1971) can be obtained as special cases of our results.

To apply our results, we view the insurer as the principal and the insured as the agent. The random state θ is interpreted as the degree of illness. There are two possible interpretations of the agent's action, a. First, a may be the level of preventive care purchased by the insured or a general decision made by him *before* the state is realized as to the amount and quality of health care to be purchased when and if he becomes ill. The second interpretation of the agent's action is as a decision made *after* becoming ill on the amount and quality of health care to purchase. These two interpretations correspond to the assumptions of Models 1 and 2, respectively. The payoff x is interpreted as the amount spent on health care.[4]

Applying the results of Section II yields the following:

i) If the insurer can observe the degree of illness of the insured θ, then from Propositions 1 and 2 the optimal insurance contract specifies a given amount to be paid by the insurer for each possible degree of illness, that is, indemnity insurance. The amount paid by the insurer is thus independent of the choice of medical care taken by the insured. The insured under this contract can choose a level of medical care which costs more or less than the amount which the insurer agrees to pay. Moreover, since the insured is obviously risk averse, Propositions 1, 2, and 3 imply that indemnity insurance is *strictly* Pareto superior to insurance based only on the cost of the medical care. All of the above results hold whether the choice of medical care is taken before or after the occurrence of an illness.

ii) If the degree of illness cannot be observed by the insurer, but the insured's choice of medical care can be observed, then from Proposition 3 there are contracts which depend on both the cost of medical care and the insured's choice of medical care which are *strictly* Pareto superior to contracts which depend only on the cost. This result holds whether the choice of medical care is taken before or after the occurrence of illness. Furthermore, when the insured chooses his level of medical care *before* the occurrence of an illness, from Proposition 4, the Pareto optimal contract (when only the total cost and the insured's choice of care are observable) stipulates that the insurer pays some share of the cost *provided that the insured chooses some prespecified level of medical care.* Otherwise, the insurer pays nothing. Even when the insured's choice of medical care is not itself

[4]Note that these interpretations of a and x require opposite assumptions on the signs of the first derivatives of U and V than were made in Section I. It is easy to check that all the results continue to hold.

directly observable, it may be optimal to employ some indirect and imperfect measure of the level of care (for example, frequency of visits to a doctor). When these measures are sufficiently precise estimates of the insured's actual choice of medical care, the optimal policy will be essentially the same as when the choice is directly observable.

C. *Compensation of Law Enforcers*

In a very interesting paper, Becker and Stigler analyze the law enforcement problem and suggest two alternative methods for improving the incentives given enforcers. Here we recast the law enforcement problem in our framework and employ the results of the previous section to obtain the optimal contract between the state and enforcers. Our results provide a firm foundation for the suggestions of Becker and Stigler. In particular, we show explicitly under what conditions each suggestion is Pareto optimal.

Becker and Stigler explore the economic incentives for malfeasance in law enforcement. They conclude that officials responsible for enforcement might lack sufficient incentives to enforce certain laws. Moreover, there may also be incentives for them to engage in malfeasance. Becker and Stigler suggest two methods for improving the incentives given enforcers. The first suggestion discourages malfeasance and lack of proper enforcement by penalizing the enforcer if such behavior is detected. The penalty is set such that it more than offsets the gain from malfeasance. This method is made operational by requiring the enforcers to "post a bond equal to the temptation of malfeasance, receive the income on the bond as long as they are employed, and have the bond returned if they behave themselves until retirement" (Becker and Stigler, p. 9). If the state detects malfeasance on the part of the enforcer, he is fired and loses the bond he posted. The second suggestion is to allow free entry into the enforcement industry. Enforcers would be rewarded based on their performance. Becker and Stigler argue that the amount of enforcement would be optimal if successful enforcers were paid the fines levied against convicted violators. These fines would equal the damages to society caused by the violator divided by the probability of conviction.

The situation described by Becker and Stigler can be stated in terms of Models 1 and 2 of Section I. The state (society) and the enforcer can be viewed as the principal and agent, respectively. The payoff x to the law enforcement activity is the revenue generated via the fine levied on convicted violators net of the costs associated with trying the accused. These latter costs include both costs borne by the accused (for example, time lost) and those borne by the state (for example, court costs). As is the case in the Becker-Stigler paper, we are not concerned with the problem of *crime prevention* or the effects of law enforcement on the level of criminal activity. In fact, we assume here that the level and type of criminal activity is completely exogenous. The schedule of fines is also taken to be exogenous. The state is interested in the degree of law enforcement as measured by the revenue generated from fine collection.

Two important aspects of the incentive problem for law enforcers may be distinguished. These aspects correspond to two distinct interpretations of the agent's action a, and the exogenous random variable θ (recall that a and θ jointly determine the payoff x). In the first, the effort of the enforcer can be interpreted as his level of investment in crime detection and apprehension capabilities and the levels of activities such as patrolling, etc. These decisions must occur *before* θ, *the level of crime activity, is known.* In this case, the problem is to provide, in an efficient contract, the proper incentives for investment in crime detection and apprehension capabilities. This problem may be analyzed using Model 1, since the agent's action (investment and patrolling) is taken before the realization of the exogenous random state (level of criminal activity).

In the second interpretation, the random state θ is the type of crimes committed, the

identities of the criminals, and other details associated with the crimes (for example, evidence, etc.). The action of the enforcer a is the effort expended in apprehending the criminals, creating cases, and the extent to which the enforcer refrains from engaging in malfeasance (for example, taking bribes, etc.). We assume for the purpose of this analysis that the enforcer knows all the details of the crimes (including the identities of the criminals) before taking his action. Society, as represented by the state, may or may not know these details. In any case, we assume contracts are agreed upon before crimes are committed. This version of the problem can be analyzed using Model 2.

Our results indicate the following:

i) With regard to the first problem, suppose the level of criminal activity can be observed *ex post*, both by the state and by enforcers. In this case, it follows from Propositions 1 and 2 that the optimal contract specifies that the state receives an amount which depends only on the level of criminal activity. In particular, the state specifies a standard for the revenue generated by enforcement activities contingent on the level of criminal activity. It then receives a certain share of this standard revenue while enforcers receive the remainder of the standard revenue plus any revenues generated in excess of the standard (or minus any shortfall of actual revenue from standard revenue). Essentially, the optimal arrangement is equivalent to one in which the state specifies a given level of investment in enforcement capability and enforcers accept full responsibility for any deviations from this level.

Regarding the second problem, when both the state and enforcers can observe the particulars of the crimes which occur, Propositions 1 and 2 imply essentially the same result but with a slightly different interpretation. Here the state specifies a total amount to be recovered by the enforcer, contingent on the particular circumstances of the crimes which have occurred. The state receives a share of this specified amount, independent of the enforcer's action. The enforcer receives the remainder of the specified amount, plus or minus any deviation of the actual amount recovered from the specified amount. All consequences of malfeasance are borne by the enforcer.

ii) If law enforcers are risk neutral, as assumed by Becker and Stigler, the Pareto optimal contract in both interpretations specifies that enforcers receive the entire payoff minus, perhaps, some constant. This follows directly from Proposition 3 and is similar to Becker and Stigler's second suggestion. In this case, enforcers purchase for a fixed amount the right to enforce laws and collect fines. There is no point in observing either the enforcer's action or the random state when the enforcers are risk neutral.

iii) If law enforcers are risk averse, contracts which depend only on the revenue generated may be inefficient. In the second interpretation, by Proposition 3, Pareto-superior results can be achieved if the particulars of the crimes committed can be observed by both parties. In this case the particular form of the Pareto optimal contract is as described in (i). Also, if there is a one-to-one relationship between the payoff and the crimes committed for any given action by the enforcer, then Pareto-superior results can be obtained if the enforcer's action can be observed.

In the first interpretation, Pareto-superior results can be achieved if either the action or the level of criminal activity can be observed by both parties. Moreover there are potential gains to monitoring (imperfectly) the activities of enforcers including detection of shirking. Under certain conditions on the monitoring technology, forcing contracts will be Pareto optimal. Under this type of contract, the enforcer would receive an amount which depends on the payoff if monitoring reveals that his action is "acceptable" (for example, there was no shirking). If his action is found to be unacceptable, he would receive a smaller amount or pay a penalty. This contract is similar to the first method suggested by Becker and Stigler, in which enforcers are required to post a bond. In order for such an arrangement to be Pareto superior to compensation

which depends only on the revenue, the state must possess means of detecting unacceptable performance by enforcers which have low probabilities of error. The restrictions on the quality of the detection mechanism needed to guarantee Pareto superiority of the Becker-Stigler proposal are quite severe.

REFERENCES

A. Alchian and H. Demsetz, "Production, Information Costs, and Economic Organization," *Amer. Econ. Rev.*, Dec. 1972, *62*, 777–95.

Kenneth J. Arrow, *Essays in the Theory of Risk Bearing*, Chicago 1970.

G. S. Becker and G. J. Stigler, "Law Enforcement, Malfeasance, and Compensation of Enforcers," *J. Legal Stud.*, Jan. 1974, *3*, 1–18.

M. Harris and A. Raviv, "Optimal Incentive Contracts with Imperfect Information," Grad. Sch. Ind. Adm., work. paper no. 70-75-76, Carnegie-Mellon Univ., April 1976.

C. M. Jensen and W. H. Meckling, "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *J. Finan. Econ.*, Oct. 1976, *3*, 305–60.

M. V. Pauly, "The Economics of Moral Hazard," *Amer. Econ. Rev.*, June 1968, *58*, 531–37.

_____, "Idemnity Insurance for Health Care Efficiency," *Econ. Bus. Bull.*, Fall 1971, *24*, 53–59.

J. G. Riley, "Competitive Signalling," *J. Econ. Theory*, Apr. 1975, *10*, 174–86.

S. A. Ross, "The Economic Theory of Agency: The Principal's Problem," *Amer. Econ. Rev. Proc.*, May 1973, *63*, 134–39.

A. Michael Spence, *Market Signaling: Information Transfer in Hiring and Related Processes*, Cambridge, Mass. 1973.

_____, "Competitive and Optimal Responses to Signals: An Analysis of Efficiency and Distribution," *J. Econ. Theory*, Mar. 1974, *8*, 1296–332.

_____, "Competition in Salaries, Credentials, and Signaling Prerequisites for Jobs," *Quart. J. Econ.*, Feb. 1976, *90*, 51–74.

_____ **and R. Zeckhauser,** "Insurance, Information, and Individual Action," *Amer. Econ. Rev. Proc.*, May 1971, *61*, 380–87.

J. E. Stiglitz, (1975a) "Incentives, Risk and Information: Notes Toward a Theory of Hierarchy," *Bell J. Econ.*, Autumn 1975, *6*, 552–79.

_____, (1975b) "The Theory of 'Screening', Education, and Distribution of Income," *Amer. Econ. Rev.*, June 1975, *65*, 283–300.

R. M. Townsend, "Efficient Contracts with Costly State Verification," Grad. Sch. Ind. Admin., work. paper no. 14-77-78, Carnegie-Mellon Univ. 1976.

R. B. Wilson, "On the Theory of Syndicates," *Econometrica*, Jan. 1968, *36*, 119–32.

_____, "The Structure of Incentives for Decentralization Under Uncertainty," in *La Decision*, Paris 1969.

R. Zeckhauser, "Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives," *J. Econ. Theory*, Mar. 1970, *2*, 10–26.