Discussion Paper No. 627

THE THEORY OF PRINCIPAL AND AGENT: PART 1.

by

Ray Rees*

October, 1984

*Department of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60201, and University College, Cardiff, CF1 1XL, United Kingdom.

The Theory of Principal and Agent

1. Introduction

A large and interesting class of problems in economics involves delegated choice: one individual has the responsibility for taking decisions supposedly in the interests of one or more others, in return for some kind of payment. Examples are a manager running a firm on behalf of its shareholders, an employee working for an employer, an accountant handling tax affairs of a client, an estate agent selling someone's house, an investment advisor administering a trust fund or share portfolio, a public policy maker, and so on. It turns out that when this situation is modelled, its formal structure is applicable to an even wider class of problems, where no formal delegation relationship is explicitly involved. For example, a person taking out fire, theft or health insurance will take a decision on the level of some activity which would reduce the risk of the event insured against and this will affect the probable income of the insurer; a firm handling dangerous chemicals will take decisions which affect the likelihood and extent of damage which would be caused to others by an accident. The theory of principal and agent is intended to apply to any situation with the following structure: one individual, called the agent and denoted A, must choose some action $\underline{a}$ from some given set of actions $\{\underline{a}\}$. The particular outcome x which results from this choice depends also on which element from some given set of states of the world, $\{\theta\}$, actually prevails at the relevant time, so that uncertainty is intrinsic to the situation. The outcome x generates utility to a second individual, the principal, denoted P. A contract is to be defined under which P makes a payment y to A. A's utility depends both on this payment y and the value of the action, $\underline{a}$. The main purpose of principal-agent theory is to

characterize the optimal forms of such contracts under various assumptions about the information P and A possess or can acquire and thereby, hopefully, to explain the characteristics of such contracts which are actually observed. It should be stressed that the term "contract" is to be interpreted very broadly. It may refer to a formal document, such as an insurance policy or sharecropping agreement, or to an implicit contract, such as may characterize an employment relationship, or to some penalty-reward system which may not formally be a contract at all--for example, the rules under which liability for damages following escape of dangerous chemicals is assessed. As usual in economics, the formal structure suggested by a particular instance of a problem is capable of much wider application.

This paper provides a survey of the literature on principal-agent theory in the following sense. I shall set out the model of the problem which has been formulated in the literature and give an exposition of the main results so far derived. This is the subject of Part 1. In Part 2 I go on to examine the main areas of application of the theory. Naturally, in developing the exposition I will refer to the papers which have developed the analysis and results. However, I make no attempt to discuss or evaluate individual papers explicitly--this is not a survey of the "who said what and when (and were they right?)" kind. The main aim of the paper is to give a clear account of the theory and some existing or potential applications, and, hopefully, to show to economists not already familiar with them that they involve some interesting and relevant economic ideas.

Part 1: Theory

1. The Formal Model

We begin by setting out the model which will be used throughout the rest

of the paper. The principal, P, has a Neumann-Morgenstern (N-M) utility function $u(x - y)$, which is not directly dependent on the state of the world, θ, and which is bounded and continuously differentiable to any required order. In particular, $u' > 0$ and $u'' \leq 0$, so we rule out risk-attracted behavior. Likewise the agent, A, has a N-M utility function $v(y,a)$ with $v_y > 0$, $v_{yy} \leq 0$, $v_a < 0$, $v_{aa} > 0$, so A, also, can only either be risk-neutral ($v_{yy} = 0$) or risk-averse ($v_{yy} < 0$). The assumption that $\underline{a}$ yields disutility to A is adopted because in most applications $\underline{a}$ is interpreted as effort or expenditure incurred by A in acting on behalf of P. Note that P is indifferent to A's choice of $\underline{a}$ as such, and cares only about the value of the outcome net of the portion of it he must pay to A. This is therefore a potent source of conflict of interest between P and A[1]. If, as we assume, A will always act in his own best interests, then in designing the contract it must be recognized that the disutility he receives from $\underline{a}$ may cause him not to act in P's best interests. Naturally we will have a lot more to say about this problem in what follows.

Without serious loss of generality, we can take the set of states of the world $\{\theta\}$ to be given by the closed unit interval $[0,1]$. A substantive assumption is that both P and A have identical probability beliefs concerning the state of the world, represented by the probability density function $f(\theta)$. This is a signficant restriction, because it might be thought that one aspect of the principal-agent relationship would be that A would possess better information on the likely occurrence of states of the world, as well as on the definition of the states themselves, than P. At some points in what follows the consequences of assuming different probability beliefs will be suggested, but the literature is entirely based on the assumption of identical probability beliefs and a full generalization is not available.

Given A's choice of $\underline{a}$, which is made before the state of the world is known, the value of the outcome x will vary with θ, and so we can write $x = x(a,\theta)$. We assume $x(\cdot,\cdot)$ is continuously differentiable to any required order, with $x_a > 0$, $x_{aa} < 0$, and, for convenience, $x_\theta > 0$, so that higher values of θ represent in some sense more favorable states. We can think of x_a as the marginal product of $\underline{a}$, and we are assuming this is always positive but nonincreasing.

With this notation, the basic principal-agent problem can be stated as follows. P is to choose a payment schedule, which in its most general form specifies a payment y to A, which could depend on x, θ, $\underline{a}$, and some other variable,[2] z, i.e., $y = y(x, \theta, a, z)$. The variable z could be thought of as something which gives (usually imperfect) information either about $\underline{a}$, or about θ, and which is costlessly available. A central assumption in principal-agent theory, which distinguishes it from the literature on incentive compatibility (for which see Hammond (1979) and the companion papers in that symposium) is that the payment schedule can depend only upon variables which both parties can observe. It is assumed that A knows $\underline{a}$ (as well as $u(\cdot)$) and can observe both x and θ. Hence different possibilities arise only in respect of the information available to P. It is always assumed that P knows $x(a,\theta)$ (as well as $v(\cdot,\cdot)$) and can always observe x. It follows that if he can observe one of $\underline{a}$ or θ he can deduce the other, ex post, from $x(a,\theta)$. We therefore have two cases of interest:

(i) P can observe $\underline{a}$ (or θ) and therefore θ (or $\underline{a}$). In that case he does not need z, since further (imperfect) information is redundant.[3] Also, the payment schedule can be taken to depend on θ alone and P chooses this payment schedule and a value of $\underline{a}$ for A in such a way as to maximize his own expected utility, subject to the constraint that A receive at least some minimum

expected utility, $\bar{v}^o$, referred to as his reservation utility.[4] As is shown in the next two sections, in this case a first-best optimum risk-sharing contract is possible, with the moral hazard or incentive problem being solved by what is known as a forcing contract. This result extends to the case in which $\underline{a}$ is observable only with some random error, provided a certain boundedness condition is met.

(ii) P can observe neither $\underline{a}$ nor θ. In this case we have a true moral hazard problem. P must recognize that given some fee schedule, A will choose $\underline{a}$ to maximize his own expected utility, and this will in general imply a value of $\underline{a}$ other than that for which the fee schedule is optimal. The lack of observability of $\underline{a}$ (and θ) means that P cannot correct this directly, and so a contraint, which we can call the incentive constraint, must be added to the reservation utility constraint in P's optimization problem.[5] In other words, P must take account of the fact that his choice of a payment schedule will determine a value of $\underline{a}$ via A's maximization procedure and thus affect the final equilibrium. In general, this leads to a departure from the optimal risk-sharing solution: there is a tradeoff between the gains from sharing risk and the need to control A's choice of $\underline{a}$--the provision of incentive. It can also be shown that when a variable z exists which gives information about $\underline{a}$, however "noisy," and which is contingent on θ, it is, except when A is risk-neutral, optimal to incorporate it into the contract and make y contingent on it, although this result would presumably be modified if z were costly to acquire.

In the following five sections we go on to analyze these cases. The analysis draws heavily on Holmström (1979) and Shavell (1979), with reliance on Harris and Raviv (1978) for rigorous proofs of what is here simply asserted. The principal-agent problem proper is essentially case (ii), but it

will be useful to consider case (i) first as a point of departure.

2. Optimal Risk Sharing

Since the central problem of principal-agent theory is to find a fee schedule which optimally trades off the benefits of risk-sharing with the costs of providing an incentive to the agent, it is useful to begin by considering the question of risk-sharing in isolation. This can be done by taking the general model just set out and fixing the value of the agent's action arbitrarily, at $\underline{a} = \underline{a}^o$. We then assume that $\underline{a}$ and/or θ can be costlessly observed so that y can be taken to depend only on θ. By a risk sharing optimum is meant a payment $y^*(\theta)$ from P to A which is Pareto efficient, i.e., which maximizes P's expected utility for some given minimum level of A's utility $\bar{v}^0$. Thus, we seek a solution to the problem:[6]

$$(R) \quad \max_{y(\theta)} \int_0^1 u(x(a^o,\theta) - y(\theta))f(\theta)d\theta \text{ s.t. } \int_0^1 v(a^o,y(\theta))f(\theta)d\theta \geq \bar{v}^o$$

The solution[7] $y^*(\theta)$, which specifies a payment from P to A at each θ, can be characterized by the following condition[8]

$$(1) \qquad -u'(x - y^*) + \lambda v_y = 0, \quad \forall\theta \in [0,1]$$

where λ can be interpreted as a conventional Lagrange multiplier which is <u>not</u>, it should be noted, a function of θ.

Since, from (1), we have that $\lambda = u'/v_y$, the ratio of marginal utilities of income of P and A at each θ, we conclude that P's nonsatiation in income implies $\lambda > 0$. It follows that the constraint in (R) must be satisfied as an equality--A receives only his reservation utility $\bar{v}^o$.

If we take two states $\theta_1 \neq \theta_2$, then (1) implies (with obvious notation)

$$(2) \qquad \frac{u'(\theta_1)}{v_y(\theta_1)} = \frac{u'(\theta_2)}{v_y(\theta_2)} \Rightarrow \frac{u'(\theta_1)}{u'(\theta_2)} = \frac{v_y(\theta_1)}{v_y(\theta_2)}$$

Thus, an implication of optimal risk-sharing is that P and A's marginal rates of substitution of income between any two states are equal. The thoroughly conventional nature of this result is brought out if we represent the situation as in figure 1. This is an Edgeworth-Bowley box, with horizontal length given by $x(a^o,\theta_1)$ and vertical length given by $x(a^o,\theta_2)$, incomes in states θ_1 and θ_2, respectively. P's indifference curves, which show loci of constant expected utility are drawn with reference to origin O_P, and A's with reference to origin O_A. The 45° lines from each origin are certainty lines; along O_PC for example, P enjoys complete income certainty. The slopes of the straight lines[9] ℓ_0, ℓ_1, ℓ_2 are the probability ratios $f(\theta_1)/f(\theta_2)$, and, given the assumption of identical probability beliefs, are the same for P and A. The indifference curve $\bar{v}^o$ corresponds to A's reservation utility, and so one equilibrium consistent with condition (1) is exemplified by point e. Clearly this is a standard type of condition for Pareto efficiency in consumption allocations, where the state-contingent incomes $y(\theta)$ and $x(\theta) - y(\theta)$ are thought of as ordinary commodities.

Two further results which turn out to be of interest in principal-agent theory can be illustrated in the figure. Suppose P is risk-neutral. Then his indifference curves <u>are</u> like the lines ℓ_0, ℓ_1, ℓ_2 in the figure.[10] Again, it follows from the assumption of identical probability beliefs that the only point at which tangency with A's reservation indifference curve $\bar{v}^o$ can take place is along his certainty line O_AC, at n. Thus, optimal risk-sharing with P risk-neutral and A risk-averse implies that P "fully insure" A by giving him a payment which is independent of θ, i.e., a <u>certain</u> income, and P bears all

the risk. The converse occurs if A is risk-neutral and P risk-averse--the equilibrium would be at m in the figure, with P receiving a guaranteed income and A bearing all the risk. With both risk-neutral, any point along the line ℓ_o is an equilibrium.

These remarks can be generalized if we enquire into the possible forms of the payments schedules which are implicitly defined by condition (1). Insight into these can be gained by differentiating the condition w.r.t. θ, recalling that $\lambda = u'/v_y$ is constant across θ. We then obtain

$$(3) \qquad -u''\left(\frac{\partial x}{\partial \theta} - \frac{dy^*}{d\theta}\right) + \lambda v_{yy} \frac{dy^*}{d\theta} = 0$$

We can now introduce the Pratt-Arrow index of absolute risk aversion, defined as:

$$r_P \equiv \frac{-u''}{u'}, \quad r_A \equiv \frac{-v_{yy}}{v_y}$$

Then, substituting for λ in (3) and rearranging gives:

$$(4) \qquad \frac{dy^*}{d\theta} = \frac{r_P}{r_P + r_A} \frac{\partial x}{\partial \theta}$$

Given risk aversion, r_P, $r_A > 0$, and so (4) implies that if, as θ increases, x increases (as was earlier assumed), then so does y, but at a slower rate. A sufficient condition for a <u>linear</u> payments schedule, i.e., a schedule of the form $y = \alpha x + \beta$, is clearly that both A and P have constant absolute risk-aversion, since in that case $r_P/(r_P + r_A)$ is constant, and integrating (4) over θ would give:

$$(5) \qquad y^*(\theta) = \alpha x(a^o,\theta) + \beta, \ \alpha = \frac{r_P}{r_P + r_A}$$

where β is a constant of integration. Moreover, if $r_P = 0$, implying that P is risk-neutral, we see immediately that we must have

$$(6) \qquad y^*(\theta) = \beta$$

implying that P bears all the risk as already illustrated in figure 1. If A is risk-neutral, $r_A = 0$ and we must have a payment schedule of the form:

$$(7) \qquad y^*(\theta) = x(a^o,\theta) - \gamma$$

i.e., A makes a fixed payment γ to P and takes the residual income.

Although the simplicity of each of these special cases is attractive, in general constant risk aversion, let alone zero risk aversion, would be regarded as rather special. If, as would more usually be assumed, r_P and r_A are both decreasing in income, the shape of $y(\theta)$ will depend on the relative changes in risk aversion as well as on the shape of $x(a^o,\theta)$ and so $y(\theta)$ could be nonlinear, convex or concave, or neither.[11] No doubt a taxonomy of cases is possible, but this would take us far from our present purpose. The case of pure risk-sharing is a preliminary step in examination of the principal-agent model, so let us return to this by considering the implications of allowing $\underline{a}$ to vary.[12]

3. The Incentive Problem

We now show that in the present case, with $\underline{a}$ observable, a "first-best" Pareto optimum with respect to both risk-sharing and A's choice of $\underline{a}$ is available. Thus, P can solve the problem

$$(FB) \qquad \max_{a,y(\theta)} \int_0^1 u(x(a,\theta) - y(\theta))f(\theta)d\theta \text{ s.t. } \int_0^1 v(a,y(\theta))f(\theta)d\theta \geq \bar{v}^o$$

with $\underline{a}$ as well as y now variable. Note that since $\underline{a}$ is to be chosen before the state of the world is known, it will not depend on θ. A first-best Pareto optimum will then be an optimal action $\underline{a}^*$ for A and an associated optimal payment schedule $y^*(\theta)$. The contract between P and A would then specify this schedule in exchange for A choosing $\underline{a}^*$. As we shall see, A does have an incentive to cheat on the contract and, given that he will receive $y^*(\theta)$, choose some $\hat{\underline{a}} \neq \underline{a}^*$. However, if P can costlessly observe $\underline{a}$ then the contract can contain a "forcing clause": if, ex post, $\hat{\underline{a}} < \underline{a}^*$ then some $\hat{y}(\theta) < y^*(\theta)$ will be paid, and of course $\hat{y}(\theta)$ can be made sufficiently unattractive as to force A to choose $\underline{a}^*$ (recall that P knows $v(y,a)$). Let us therefore examine the first-best solution corresponding to the absence of the incentive problem.

The solution to FB can be characterized by the conditions:

$$(8) \qquad -u' + \lambda v_y = 0$$

$$(9) \qquad E[u' x_a + \lambda v_a] = 0$$

where the expectations operator E has replaced the integral notation. Again, the Lagrange multiplier $\lambda > 0$, given $u' > 0$. Thus A receives only $\bar{v}^o$. Note that, since $\underline{a}$ is chosen optimally, condition (8) is identical to (1), and we have optimal risk-sharing just as before: given choice of $\underline{a}$, P and A share the risk associated with the resulting distribution of x in a Pareto-efficient way. The new element is condition (9), which relates to the optimal choice of $\underline{a}$ and has a straightforward interpretation. In any one state of the world,

$u'x_a$ can be interpreted as the marginal value product of $\underline{a}$ measured in terms of P's utility or "u-utils," i.e., $\frac{du}{da} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial a}$. Then, λv_a can be interpreted as the marginal cost of $\underline{a}$ in "u-utils": at the optimum, $\lambda = u'/v_y$ gives the number of "u-utils" P has to give up to yield A one "v-util"; while v_a gives the number of "v-utils" A requires to be paid to supply the marginal bit of $\underline{a}$ (recall $v_a < 0$). Thus $(u'x_a + \lambda v_a)$ is net marginal value product of $\underline{a}$ in "u-utils"[13]. Now if $\underline{a}$ were state-contingent P would choose $\underline{a}$ so as to set this net marginal value product at zero (marginal value product equals marginal cost) in each state. But because $\underline{a}$ must be chosen before the state of the world is known, the marginal value product and marginal cost are equalized in expected value terms--on average across all states.

Since our earlier analysis of the form of the risk-sharing contract was conducted for arbitrary $\underline{a}$, it applies equally now for $\underline{a}$ at the optimal value of $\underline{a}^*$. The important point is that observability of $\underline{a}$ implies that Pareto-efficient risk-sharing is still possible. The two special cases of P risk-neutral and A risk-neutral are again of interest. Thus, suppose P is risk-neutral so that u' is a constant. Then from the earlier analysis we know that y^* is a constant and so, since $\underline{a}$ is independent of θ, we must have that in each state $v(y^*,a^*) = \bar{v}^o$. On standard assumptions, we can draw this contour of A's utility function as $\bar{v}^o$ in figure 2, treating y as certain. If we treat u' as a constant in (9), substitute for λ, and note that both v_a and v_y are independent of θ, we have:

$$E[x_a] = \frac{-v_a}{v_y} \tag{10}$$

or, at the optimum the expected marginal product of $\underline{a}$ is equated to A's unique marginal rate of substitution between $\underline{a}$ and income. Then, define the

function:

$$(11) \qquad \bar{x}(a) = \int_0^1 x(a,\theta)f(\theta)d\theta$$

which gives the expected value across θ of x at each $\underline{a}$. This is graphed as the curve $\bar{x}$ in figure 2. Then the optimum for a risk-neutral P is given by $\underline{a}^*$ in the figure, since at this point the slope of $\bar{x}(a)$ is equal to the slope v_a/v_y of $\bar{v}^o$. This has a straightforward interpretation. To induce A to choose any given $\underline{a}$, P will offer him a fixed payment (efficient risk-sharing) which must lie on $\bar{v}^o$ ($\lambda > 0$, so A receives only his reservation utility). Hence, $\bar{v}^o$ is a type of "total cost curve" to P.[14] Since P is risk neutral, the output distribution $x(a,\theta)$ can be valued at its expected value, and so the vertical distance between the two curves in figure 2 can be thought of as P's "expected net income." P then seeks to maximize this, implying that he wants A to choose $\underline{a}^*$, and pays him y^* in exchange. P's income distribution, $x(a^*,\theta) - y^*$, will then be given from the distribution $x(a^*,\theta)$ of which $\bar{x}^*$ is the expected value.

In the case where A is risk-neutral, P retains a constant payment, γ, and so again u' is constant. But v_y is also constant (risk-neutrality), and so (9) becomes:

$$(12) \qquad E[x_a] = -E[\frac{v_a}{v_y}]$$

In this case, the optimal $\underline{a}$ equates the expected marginal product with the expected value of A's marginal rate of substitution between $\underline{a}$ and income, which, for $\underline{a} = \underline{a}^*$, varies with $x(a^*,\theta) - \gamma$

Thus, when P can observe $\underline{a}$ or θ costlessly, the first-best is

attainable. If he can observe neither, then the nature of the incentive or moral hazard problem is as follows. Since θ cannot be observed the payment must be expressed as conditional on x. If P naively seeks to implement the solution derived in this section, he could find $x = x(a^*,\theta)$, and offer A the payment schedule $y^*[x(a^*,\theta)]$, that is, he rewards A upon the occurrence of an observed x on the assumption that $\underline{a} = \underline{a}^*$ and that the observed x is derived from the distribution $x(a^*,\theta)$. If A is individually rational, he will then solve the problem:

$$\text{(AR)} \qquad \max_a \int_0^1 v(y^*[x(a,\theta)],\ a)f(\theta)d\theta$$

That is, he will choose an $\underline{a}$ in the light of the income distribution which will result under the payment schedule $y^*(x)$. But there is no guarantee in general that the solution to (AR), which we can denote by $\hat{a}$, is the same as a^*, the solution to (FB). For example, if P is risk-neutral $y^*(s) = \beta$. But substituting into v in (AR) will result[15] in $\hat{\underline{a}} = 0$: why should A incur <u>any</u> disutility if he will be paid anyhow? More generally, the solution to (AR) must satisfy the condition:

$$\text{(13)} \qquad E[v_y(\frac{dy^*}{dx}\, x_a + \frac{v_a}{v_y})] = 0$$

This can be compared to the condition determining $\underline{a}^*$ which, from (8) and (9) is

$$\text{(14)} \qquad E[u'(x_a + \frac{v_a}{v_y})] = 0$$

In income terms, the marginal product of $\underline{a}$ to A is $\frac{dy^*}{dx}\, x_a$, since the effect on

his own income of a change in x is determined via the payment schedule, while to P the marginal product of $\underline{a}$ is x_a, given y. Since, from (4) $dy^*/dx < 1$ at a^*, the two differ in their valuation of the marginal product of $\underline{a}$ quite apart from the differences in their marginal utilities of income.

Intuitively, the problem is that since P's choice of $\underline{a}$ is not optimal for A given the associated payment schedule $y^*(x)$, it will be possible for A to make himself better off by choosing $a \neq a^*$ if he can do this unobserved and unpenalized. This is the moral hazard problem. Before examining how P must deal with this, we consider two cases in which the incentive problem does not arise. The first is that in which A is risk-neutral. Here the first-best solution is available essentially because when P gives A the first-best payment schedule, A's optimal choice of $\underline{a}$ is the first-best level of $\underline{a}$, so the incentive constraint is in effect not binding. The second case is where, although $\underline{a}$ cannot be observed perfectly, it can be observed with a random error which is independent of θ. In this case, by means of a forcing contract, the first best is again available.

A is Risk-Neutral

Harris and Raviv (1978) and Shavell (1979) show that if A is risk neutral, so that v_y is a constant, then P can achieve a first-best allocation and no incentive problem arises. This can be expressed in the form of the proposition that if A is risk neutral, a contract which specifies y contingent only on x is at least as good as one which makes y contingent on $\underline{a}$ and θ as well as x. Thus, information about $\underline{a}$ or θ has no value, or, to put this another way, it does not matter if $\underline{a}$ and θ cannot be observed. Here we will give a simple account of the proposition which brings out its essential point.

Recall that first-best risk-sharing when A is risk-neutral requires that P retain a fixed payment γ and A receive the residual uncertain income

$x(a,\theta) - \gamma$. The condition for first-best optimal choice of $\underline{a}$, from (12) is

$$E[x_a] = - \frac{E[v_a]}{v_y}$$

If now, even though $\underline{a}$ is not observable, P offers A the same payment schedule (i.e., asks for the same fixed payment γ) then A will choose $\underline{a}$ to solve

$$\text{(ARN)} \qquad \max_a \int_0^1 v(x(a,\theta) - \gamma, a) f(\theta) d\theta$$

which, with v_y constant, yields precisely condition (12). Thus, in this case, A's choice of $\underline{a}$ does not differ from P's given the payment schedule. A will of course accept the fee schedule $x(a,\theta) - \gamma$, because, since γ is derived from the solution to the first-best problem, it satisfies the reservation utility constraint. Essentially then, the incentive constraint described previously is non-binding at P's optimum.

4. Imperfectly Observable $\underline{a}$

A relaxation of the assumption of nonobservable $\underline{a}$, which turns out to have strong implications, was suggested by Harris and Raviv (1976, 1978).[16] Suppose that P can observe a random variable $\alpha = \underline{a} + \varepsilon$, where ε has zero mean and probability $\phi(\varepsilon) > 0$ on some interval $[\varepsilon_0, \varepsilon_1]$ and zero elsewhere. Thus there is a kind of measurement error in P's observation of $\underline{a}$. The key point is that ε is independent of θ, the state of the world. Then it is easy to show that P can adopt a forcing contract to achieve the first-best solution, and so a moral hazard problem does not really arise. Suppose for example that ε is uniformly distributed over $[\varepsilon_0, \varepsilon_1]$, as illustrated in figure 3, where a^* is again P's first-best values of a. It is of course assumed that P knows the

function $\phi(\varepsilon)$. Then P need simply threaten an arbitrarily low y[17] if he observes some $\alpha < \underline{a}^* + \varepsilon_0$, since this occurs if and only if $\underline{a} < \underline{a}^*$. Since A will not choose $\underline{a} > \underline{a}^*$, for a given payment schedule, he will then choose $\underline{a}^*$.

If $\phi(\varepsilon)$ was not positive only on an interval, i.e., if it were positive everywhere on the real line (for example, if ϕ is the normal distribution), then P is involved in a problem of hypothesis testing. At its simplest, this would involve choosing some critical value, α^*, such that an observation $\alpha < \alpha^*$ would be taken to indicate $a < a^*$ even though there is some positive probability that $a = a^*$. Hence P would have to weigh up the losses from "type 1" and "type 2" errors--respectively, the errors of falsely rejecting $a = a^*$ and of falsely accepting that--in choosing α^*. This problem does not appear to have received explicit attention in the literature, possibly because in outline it looks less interesting than the case in which P observes not a simple distorted value, but rather a variable z which depends upon both $\underline{a}$ and θ. The implication of such a possibility of observation will be considered in section 6, below. First, we consider the solution to the mixed risk-sharing and incentive problem.

5. Solutions to the Principal-Agent Problem

We now assume that P can observe only the outcome x and has no information whatever about $\underline{a}$ and θ. Then his problem is taken to be that of choosing a payment schedule $y^*(x)$ which maximizes his expected utility, taking into account the constraints that A must receive at least his reservation utility and will, given any y(x), choose an $\underline{a}$ which maximizes his own expected utility.

A formal approach to this problem would be to take the problem FB and append to it condition (13) as a constraint, implying that P's choice of y(x) will now take account of its affect on A's choice of $\underline{a}$ via A's maximization

condition. This indeed was the approach adopted by Harris and Raviv (1976), Ross (1973), and Spence and Zeckhauser (1971). It turns out, however, that this problem is not well-behaved. If $y(x)$ is not restricted to some finite interval at each x, an optimal solution to the problem may well not exist, as shown by Mirrlees (1975). If $y(x)$ is restricted to a finite interval, as is quite reasonable, the derivative $y'(x)$ which appears in condition (13) may not in fact exist at all points. Since the approach to solution of the problem, based on the calculus of variations, takes $y'(x)$ as a control variable in solving the problem, this is a serious weakness.

An alternative approach, suggested by Mirrlees (1974, 1975), and developed further by Holmström (1978) gets around this difficulty by eliminating θ from the problem and regarding x itself as the basic random variable with respect to which expected values are taken. Thus, given some $\underline{a}$, there is an x for each θ with associated probability density $f(\theta)$. Then the function $x(a,\theta)$ and the density function $f(\theta)$ jointly determine a probability distribution for x. An increase in a is taken to shift this distribution rightward, with the proviso, required on technical grounds, that the upper and lower bounds of its distribution, which will be at $x_1 \equiv x(a,1)$ and $x_0 \equiv x(a,0)$ respectively, are invariant to changes in $\underline{a}$. This means that however much $\underline{a}$ the agent chooses, the outcome x is unchanged in the most favorable state $\theta = 1$, and the least favorable state $\theta = 0$.

It turns out, however, that this approach does not guarantee the uniqueness of a solution to condition (13), i.e., there may be multiple solutions to A's problem of maximizing his expected utility subject to a given payment schedule $y(x)$, and this may in turn imply that the conditions for a solution to the principal-agent problem derived under the Mirrlees-Holmström procedure do not in fact characterize an optimum. This is neatly

illustrated[18] by a diagram in Grossman and Hart (1983), and reproduced here as figure 4. In the figure, η refers to a payment schedule (not a value of y) ranked in (continuous) order of P's preference from left to right, and $\underline{a}$ is again A's action. The possibility of multiple choice of $\underline{a}$ for given payment schedule η is reflected in the shape of the curve $a(\eta)$, which illustrates A's choice of $\underline{a}$ for each η. However, given any η, A will choose only an $\underline{a}$ on the dotted portion of the curve because he prefers less $\underline{a}$ to more and so these points dominate the others. P's indifference curves are as shown (though $\underline{a}$ is not a direct argument of u it enters indirectly via x). Then, the Mirrlees-Holmström procedure characterizes P's optimum as being at point E in the figure, since it yields the highest utility of all the points which satisfy A's first-order condition (13), but T is in fact the true optimum, since it is the best point for P out of the points which he can actually induce A to choose--setting $\hat{\eta}$ would induce point W, not E. The existence of this possibility is a pity[19] since, as Holmström shows, the procedure he adopts leads to a relatively simple characterization of an optimal solution to the principal-agent problem.

It would seem from this discussion that only two courses are open. One could guarantee a well-behaved problem by making some more-or-less drastic simplification of the structure.[20] Alternatively, one could assume the uniqueness problem does not exist and just enjoy the niceness of the results that follow. In a certain sense the problem is purely a theoretical one: if P knows $v(y,a)$ and $x(a,\theta)$, then he knows the relationship between A's choice of $\underline{a}$ and any payment schedule that might be chosen, and so, in principle at least, could always find a global optimum. For example, in figure 4, if P knows the curve $a(\eta)$, then why should he be fooled into choosing point E? However, given our analytical concerns the problem is a substantive one: we

wish to characterize an optimal solution using standard procedures and have to take seriously the risk that they do not work properly for all cases.

Here we provide an exposition of the Holmström-Mirrlees approach, since, taken across the literature, this seems to combine the most general form of problem-structure with the simplest statement of the results, one which gives a clear insight into the effects of introducing the incentive constraint. Thus:

(i) $v(y,a) \equiv v_1(y) - v_2(a)$, the additive separability assumption.

(ii) Take x as the random variable, whose distribution is derived from $x(a,\theta)$ and $f(\theta)$, and is written $\phi(x,a)$. It remains the case that the payment schedule is expressed in terms of the observable variable, x. However, x itself is now invariant to $\underline{a}$, since it is in essence the state variable.

(iii) An important property of ϕ is: $\phi(x,a) \equiv 0$ for $x \notin [x_0,x_1]$, for all $\underline{a}$, and $\phi(x,a) > 0$ for $x \in [x_0,x_1]$.

Figure 5 illustrates the assumed type of behavior of ϕ as $\underline{a}$ changes. For higher $\underline{a}$ the whole distribution shifts to the right, but with unchanged support, x_0, x_1. Note that, for given x, it is assumed:

(iv) The derivatives ϕ_a, ϕ_{aa} are well-defined, with $\phi_a \lessgtr 0$, as the figure illustrates. Thus, increased $\underline{a}$ makes low values of x less, and high values of x more probable.

(v) The distribution resulting from a higher value of $\underline{a}$ is always preferred by P to one resulting from a lower value of $\underline{a}$. Thus increasing $\underline{a}$ leads to "better" distributions of x.

The incentive constraint now is the first-order condition for solution of the problem of maximizing A's expected utility w.r.t. $\underline{a}$, i.e.,

$$(A) \qquad \max_{a} \int_{x_0}^{x_1} v_1[\hat{y}(x)]\phi(x,a)dx - v_2(a)$$

yielding:[21]

$$(15) \qquad \int_{x_0}^{x_1} v_1[\hat{y}(x)]\phi_a(x,a)dx - v_2'(a) = 0$$

where $\hat{y}(x)$ is any given payment schedule. Given (15), P now has the problem of finding a function $y(x)$ to solve

$$(PA) \qquad \max_{y(x),a} \int_{x_0}^{x_1} u(x - y(x))\phi(x,a)dx$$

$$\text{s.t.} \int_{x_0}^{x_1} v_1[y(x)]\phi(x,a)dx - v_2(a) \geq \bar{v}^o$$

$$v_2'(a) = \int_{x_0}^{x_1} v_1[y(x)]\phi_a(x,a)dx$$

where the first constraint is again A's reservation utility and the second is the incentive constraint from (15). It should be recalled that x is <u>not</u> a variable in this optimization--it plays the same role as did θ in the earlier problem. Then, associating multipliers λ and μ (not dependent on x) with the respective constraints we have the conditions:[22]

$$(16) \qquad \{-u' + \lambda v_1'\}\phi(x,a) + \mu v_1'\phi_a(x,a) = 0$$

$$(17) \qquad \int_{x_0}^{x_1} u\phi_a dx + \lambda[\int_{x_0}^{x_1} v_1\phi_a dx - v_2'] + \mu\{\int_{x_0}^{x_1} v_1\phi_{aa}dx - v_2''\} = 0$$

First note that in (17), the middle term in square brackets is zero from the incentive constraint. The conditions can then be written:

$$(18) \qquad \frac{u'}{v_1'} = \lambda + \mu\phi_a/\phi$$

$$(19) \qquad E[u\phi_a] = -\mu E[d^2v/da^2]$$

where $E[d^2v/da^2] \equiv \int_{x_0}^{x_1} v_1\phi_{aa}dx - v_2''$ is a notation designed to bring out the fact that the incentive constraint is of the form $E[dv/da] = 0$, and the third term in (17) is then simply the derivative of this w.r.t. a. It can be shown (see Holmström, p. 90), that $\mu > 0$, so the incentive condition represents a binding constraint[23] on P. (18) then shows that risk-sharing will not be Pareto-efficient (compare condition (1)) and is distorted by the need to take account of the incentive effects on A, i.e., the effect of the choice of y, given x, on A's choice of a and hence the effect on the probability of getting x, ϕ_a. The simplicity of the earlier results on the form of the risk-sharing contract is also lost: this now cannot be completely determined by the attitudes to risk, but will also depend in general on how ϕ_a and ϕ vary with x, i.e., on the underlying functions $f(\theta)$ and $x(a,\theta)$.[24]

However, Holmström is able to establish some interesting results on the precise way in which the payment schedule will change as a result of the incentive constraint, and these can be described quite simply if we redefine the problem slightly. Interpret λ now <u>not</u> as the multiplier associated with the reservation utility constraint, but simply as a <u>fixed weight</u> given to A's expected utility[25] in forming the maximand of the problem:

$$(PA') \qquad \max_{y(x),a} \int_{x_0}^{x_1} u\phi dx + \lambda \int_{x_0}^{x_1} \{v_1 - v_2\}\phi dx \text{ s.t. } \int_{x_0}^{x_1} \frac{dv}{da}\phi dx = 0$$

The solution to this problem is evidently identical in form to conditions (18)

and (19), but we have the added advantage that λ is now a constant (and is certainly non-zero--it is not clear from Holmström's analysis that $\lambda \neq 0$ in general, since having to meet the incentive constraint could lead to A receiving more than $\bar{v}^o$). Then consider condition (18), and note that, because of diminishing marginal utility, $u'(x - y)/v_1'(y)$ is increasing in y for <u>given</u> x. Now suppose, for given x, we have the <u>first-best</u> $y^*(x)$ such that:

$$(20) \qquad \frac{u'(x - y^*(x))}{v_1'(y^*(x))} = \lambda$$

There are two sets of values of x of interest: first, $X^+ = \{x | \phi_a(x,a) > 0\}$ and secondly, $X^- = \{x | \phi_a(x,a) < 0\}$ (refer back to figure 5). Then if we add to (20) the term $\mu\phi_a/\phi$, with λ constant, to obtain (18), we observe, since $\mu > 0$, that u'/v_1' must increase when $\phi_a > 0$, and decrease when $\phi_a < 0$. That is, $y(x) > y^*(x)$ when $x \in X^+$, and $y(x) < y^*$ when $x \in X^-$. Thus the incentive effect requires deviation from optimal risk-sharing by increasing A's payment in states when increased $\underline{a}$ increases their probability, and by reducing A's share in states when increased $\underline{a}$ reduces their probability. One implication of this is that a risk-neutral P would not now make a fixed payment to A.

The second-best solution is <u>strictly</u> worse for both P and A than the first-best, implying that there are efficiency gains to be had if only $\underline{a}$ could be observed[26] by P. This then leads to the question: supppose some variable z can be acquired costlessly by P, which gives some kind of information about $\underline{a}$. Should it then be incorporated into the contract, in the sense that payment to A should be contingent on observed z, so that y would differ, for <u>given</u> x, if z differed? As the next section shows, the answer is yes, in general, even though z may give very imperfect information about $\underline{a}$.

6. The Use of Imperfect Information About $\underline{a}$

Suppose now that although $\underline{a}$ and θ cannot be observed directly, there is some variable z which provides information about a in the following specific sense. The value of z depends on $\underline{a}$ and θ, i.e., we can write $z(a,\theta)$, so that a change in a shifts the entire distribution of z. Then, since we have $x(a,\theta)$, there will, for some given $\underline{a}$, be a joint probability distribution of x and z. P is assumed to be able to observe z costlessly, and also to know the joint probability density function, which can be written $\phi(x,z,a)$. Then, the question is, given some outcome for x, will it pay P to use the outcome value of z in determining the payment he makes to A?

On the face of it, it is not immediately obvious that information of this kind necessarily would be incorporated into a contract by making y contingent on both x and z. As Harris and Raviv argue, although the increased information about probable values of $\underline{a}$ is a benefit, there is also a cost in that this information is uncertain, and so if P and A are risk-averse this may make the incorporation of z into the contract unattractive. However, as Harris and Raviv, Shavell and Holmström show, it is <u>always</u> optimal to incorporate such information into the contract when $\underline{a}$ and θ are not observable (except, as we have already indicated, where A is risk-neutral), so that the benefit of extra information outweighs whatever cost the extra uncertainty might impose (although the assumption that z can be observed costlessly is important here).

For a proof of this proposition the reader is referred to Shavell (1979, appendix, p. 69). Here we adopt the simpler and more transparent approach of Holmström, who incorporates z into P's optimization problem in a straightforward way and then shows how condition (18) is affected.

The problem can now be taken as that of defining a payment schedule

$y(x,z)$ where, formally, z, like x, is treated as a state variable. The function $\phi(x,z,a)$ gives the joint probability of x and z given $\underline{a}$, and P's problem now becomes

$$(PAz) \qquad \max_{y(x,z),a} \int_{x_0}^{x_1} \int_{z_0}^{z_1} u(x - y(x,z))\phi(x,z,a)\, dzdx$$

$$\text{s.t.} \quad \int_{x_0}^{x_1} \int_{z_0}^{z_1} v_1(y)\phi(x,z,a)\, dzdx - v(a) \geq \bar{v}^o$$

$$\int_{x_0}^{x_1} \int_{z_0}^{z_1} v_1(y)\phi_a(x,z,a)\, dzdx - v'(a) = 0$$

which differs from the previous formulation (PA) only in that expectations must be taken with respect to the joint distribution of the random variables x and z. Since we maximize w.r.t. y at each pair of values (x,z) we obtain as the counterpart to (18)"

$$(21) \qquad \frac{u'}{v_1'} = \lambda + \mu \frac{\phi_a(x,z,a)}{\phi(x,z,a)}$$

Thus, if ϕ_a/ϕ varies with z, the payment P will make to A on observation of x will now be modified in the light of the observation of z. For example, if ϕ_a varies inversely with z, the payment P makes to A for a given x will be lower when z is incorporated into the contract than when it is not. The essential reason for incorporating z into the contract is not that it provides additional information about the likely value of $\underline{a}$--after all, given the payment schedule, P knows exactly what that will be--but rather because it provides a more discriminating way of giving A an incentive to increase his value of $\underline{a}$. Or, equivalently, it reduces the cost to P of providing A with the right kind of incentive. This can be put intuitively in the following

way. If the contract depends on x alone, then, given the distribution $\phi(x,a)$ a high value of x could be observed, and a correspondingly high payment made to A, with given probability, even though $\underline{a}$ is low. Similarly, a low value of x could be observed and a low payment made even when A chooses a high $\underline{a}$. Each of these possibilities is undesirable from the point of view of providing an incentive to A to choose high $\underline{a}$. If now some variable z is observable, whose value, let us assume, also increases with $\underline{a}$ and θ, it becomes less likely that high values of both x and z would be observed when $\underline{a}$ is in fact low, and also less likely that low values of both x and z would be observed when $\underline{a}$ is in fact high. Thus, incorporating z into the contract reduces the chance of wrongly rewarding low $\underline{a}$ and wrongly penalizing high $\underline{a}$. This improves the incentive properties of the contract.[27]

7. Conclusions to Part 1

In this part of the paper we have set out what may be regarded as the "basic" principal-agent model and have explained the main results on contracts which have been derived from this. Where $\underline{a}$ is either directly observable or observable up to a random error, a first-best risk-sharing contract is feasible, which will involve a clause penalizing A for choosing an $\underline{a}$ below the optimal level. In this case if A is risk-neutral P retains a fixed sum and A bears all the risk; if P is risk-neutral A receives a fixed sum and P bears all the risk. Indeed, if A is risk-neutral then the first best is available to P even when he cannot observe $\underline{a}$, since A, in acting in his own interests, will choose the first best value of $\underline{a}$ provided P offers him the first-best fixed payment. The incentive problem proper then arises when A is risk averse and neither $\underline{a}$ nor θ is observable. In that case there is a genuine second-best problem. The optimal contract now must take account of the need to influence A's choice of $\underline{a}$--the incentive requirement--and so will have to

provide for a different payment schedule than that which optimally shares risk. For example, a risk-neutral P would not now pay A a fixed sum. In general, the incentive requirement calls for a higher payment to A at relatively high values of x and a lower payment at relatively low values of x, as compared to optimal risk-sharing, in order to induce A to increase $\underline{a}$ from the below-optimal level which his distaste for $\underline{a}$ would otherwise lead him to choose. Finally, if there is costlessly available information on some variable z whose distribution of values depends on $\underline{a}$, the optimal second-best contract will always incorporate this to make the payment to A contingent on both x and z, essentially because it reduces the chances of wrongly rewarding A for low $\underline{a}$, and wrongly penalizing him for high $\underline{a}$, and thus improves the incentive properties of the contract.

There have been a number of interesting extensions to the "basic" model in the recent journal literature, and there are also some possible extensions, not yet made, which could be discussed. However, it would seem most useful to consider these at the conclusion of this paper, after we have examined some applications of the basic model. This is the subject of Part 2.

Notes

Part 1

[1]The two papers by Ross (1973, 1974), which initiated study of the principal-agent problem, in fact assume that a does not enter into A's utility function. A conflict of interest then arises if the two utility functions are essentially different, i.e., if there do not exist $\alpha > 0$, β such that $v \equiv \alpha u + \beta$ (recall that N-M utility functions are unique up to a positive linear transformation). If a is excluded from v, however, the model is then one purely of risk-sharing, and does not encompass the problem of incentives and moral hazard, which would generally be regarded as a central issue in the principal-agent relationship. The following sections should make this point clear.

[2]Needless to say, in a fully general treatment any or all of x, a and z could be vectors rather than scalars. Nothing essential is lost by restricting the exposition to the simpler case considered here.

[3]Harris and Raviv (1978) provide rigorous proofs of this and some other assertions which will be taken for granted here. Harris and Raviv also show that the results given in this paper for the case in which a is chosen before the state of the world is known extend quite readily to the case where a is chosen after θ is known.

[4]This reservation utility is never discussed at any length in the literature. It is usually taken as "market determined" and left at that. Yet it turns out to be important because at the solution to most models A receives only $\bar{v}^o$, with P appropriating all the gains from trade (indeed, the only paper to consider the issue of whether, in equilibrium, $v > \bar{v}^o$, explicitly, is Grossman and Hart (1983), in Proposition 3). Clearly what is required is some theory of the market interaction between principals and agents, a suggestion for further research made at the very outset, by Ross (1973), and so far not taken up.

[5]This additional constraint then creates a second-best problem relative to the problem in case (i). In fact the structure is quite analogous to the types of problems first considered in the theory of the second-best--see, for

example, Lipsey and Lancaster (1956), and Davis and Whinston (1965).

[6]Strictly we should also impose upon this problem the condition that $y(\theta) \in [y^o, x(a^o,\theta)]$, $\forall\theta$, where $y^o \gtrless 0$ is some lower income bound for A, but for simplicity we assume an interior solution: in each state both P and A receive a positive share of the outcome x.

[7]This is found by forming the function $\{u + \lambda(v - v^o)\}f(\theta)$ and finding the maximum of this w.r.t. y at each θ; i.e., we carry out a pointwise maximization.

[8]This solution appears first to have been developed by Borch (1962).

[9]It is no accident that the convex indifference curves are tangent to these lines along the certainty lines O_PC and O_AC. For example, since expected utility is constant along an indifference curve, we must have

$$\frac{dy(\theta_2)}{dy(\theta_1)} = \frac{-f(\theta_1)v_y(\theta_1)}{f(\theta_2)v_y(\theta_2)}.$$

But at $y(\theta_1) = y(\theta_2)$, $v_y(\theta_1) = v_y(\theta_2)$, and so marginal rate of substitution of state-contingent incomes is equal to the probability ratio at such a point.

[10]Thus, in footnote 9, set $v_y(\theta_1) = v_y(\theta_2)$ for all y, since risk-neutrality implies that A's marginal utility of income is independent of income. Then $dy_2/dy_1 = -f(\theta_1)/f(\theta_2)$ for all y. Then the same would obviously hold for P's marginal rate of substitution.

[11]Thus, differentiating (4) w.r.t. θ would show that even if $d^2x/d\theta^2$ is signed, and both utility functions have diminishing risk aversion, we cannot in general sign $d^2y^*/d\theta^2$—everything depends on the relative rates at which risk aversion of A and P diminish.

[12]The effect of divergent probability beliefs of P and A can be briefly indicated. In that case (1) becomes

$$(1') \qquad -u'f(\theta) + \lambda v_y g(\theta) = 0$$

where $g(\theta) \neq f(\theta)$ is A's probability density on θ. (4) then becomes

$$(4') \qquad \frac{dy^*}{d\theta} = \left(\frac{r_P}{r_P + r_A}\right)\frac{\partial x}{\partial \theta} + \left(\frac{1}{r_P + r_A}\right)\left\{\frac{g'}{g} - \frac{f'}{f}\right\}$$

Thus the way in which the optimal payment y^* varies with x or θ now depends also on the probability beliefs of P and A, and constant absolute risk aversion is no longer sufficient for linearity of the payments schedule, and risk-neutrality of either A or P no longer gives the simple "full insurance" results previously described.

[13]Thus, the dimension of λ is

$$\frac{\text{u-utils/\$}}{\text{v-utils/\$}} = \frac{\text{u-utils}}{\text{v-utils}}$$

and the dimension of λv_a is

$$\frac{\text{u-utils}}{\text{v-utils}} \cdot \frac{\text{v-utils}}{\text{units of } \underline{a}} = \frac{\text{u-utils}}{\text{units of } \underline{a}}.$$

[14]Grossman and Hart (1983) use the idea of the cost of inducing A to choose some $\underline{a}$ extensively to develop a wide set of interesting results for a special case of the principal-agent problem, in which A's utility function takes the form $v(y,a) \equiv G(a) + K(a)V(y)$, encompassing both additive ($K(a) \equiv 1$) and multiplicative ($G(a) \equiv 0$) separability, and where the action $\underline{a}$ affects the probabilities of occurrence of a fixed, finite set of outcomes rather than the values of the outcomes themselves.

[15]In this case strictly the nonnegativity condition $a \geq 0$ would have to be added to AR, otherwise $v_a < 0$ implies $\hat{a} \to -\infty$.

[16]See also the comments by Shavell (1978), fn. 4, and Holmström (1978), p. 75.

[17]For example, a "Stalinist" solution might have A taken out and shot.

[18]From a suggestion by Andreu Mas-Colell. The recognition of the non-uniqueness problem itself is attributed to an unpublished paper, Mirrlees (1975).

[19]The non-uniqueness problem is not hard to establish. To guarantee a unique global solution to the problem max $J(a) \equiv \int_0^1 v(y[x(a,\theta)], a)\, f(\theta)d\theta$ we require J to be strictly concave in $\underline{a}$ for all $\underline{a}$ but:

$$J''(a) = \int_0^1 \{y'x_a[v_{yy}y'x_a + 2v_{ya} + \frac{v_{aa}}{y'x_a}] + v_y x_a^2 y'' + v_y y' x_{aa}\} f(\theta)d\theta$$

which cannot be signed in general because the sign of y'' is not known. Certainly multiple local optima cannot be ruled out and Mirrlees' example shows that they can plausibily exist.

[20]Thus Grossman and Hart take a special form of the v-function, assume P is risk-neutral (although most of their results generalize), take a finite set of outcomes $\{x_1,\ldots,x_n\}$ independent of $\underline{a}$, and make the associated probabilities $\{f_1,\ldots,f_n\}$ functions of $\underline{a}$. Holmström assumes the v-function is additively separable in y and $\underline{a}$ and takes a fixed interval of x-values over which the probability distribution changes with $\underline{a}$, though, as already noted, this is still not sufficient to guarantee uniqueness.

[21]The usefulness of the separability condition on v can be seen in the simplicity of condition (15).

[22]Strictly we should take account of the constraint that y(x) lies in some closed interval at each x, e.g., $0 \leq y \leq x$. For simplicity, I assume here that the solution is always at an interior point of such an interval. The conditions are obtained by differentiating through by y for each x, and differentiating through by $\underline{a}$ for all x (since a is chosen before x is known, y after).

[23]It is easy to verify for Holmström's model that, as we saw earlier, if P solves his first-best problem then the incentive constraint will not be satisfied in general. Thus, in the present case we would have

$$\max_{y(x),a} \int_{x_0}^{x_1} u(x - y(x))\phi(x,a)dx \quad \text{s.t.} \int_{x_0}^{x_1} v_1[y(x)]\phi(x,a)dx - v_2(a) \geq \bar{v}^o$$

yielding

$$u'/v_1' = \lambda \quad \text{and} \quad E[u\phi_a] + \lambda\{E[v_1\phi_a] - v_2'\} = 0$$

where the second condition clearly differs from the incentive constraint.

However, $\mu > 0$ is a rather stronger result than $\mu \neq 0$, and in his proof of this Holmström assumes the second-order condition for problem (A) is satisfied, something which may not be true in general. Also note that where A is risk neutral, in effect $\mu = 0$ because the incentive constraint is not binding.

[24]Thus, differentiating through (18) gives:

$$\frac{dy^*}{dx} = \frac{r_P}{(r_P + r_A)} \frac{u'}{v_1'} + \frac{\mu}{r_A + r_P} \left\{ \frac{\phi_{ax}}{\phi} - \frac{\phi_a \phi_x}{\phi^2} \right\}$$

So, for $\mu > 0$, $r_P = 0$ does not imply $dy^*/dx = 0$ unless restrictions are placed on ϕ. This expression may perhaps explain why Grossman and Hart found it not possible to assign even such simple properties as monotonicity to $y(x)$.

[25]In other words, we now simply seek a Pareto optimum relative to the incentive constraint, where λ determines the utility distribution in the final allocation. This latter may, but need not, coincide with $\bar{v}^o$ for A. This problem could have the interpretation that P and A bargain efficiently over the contract, and P does not necessarily get all the gains from trade.

[26]Note that A would <u>also</u> benefit from a move to the first-best. This might suggest the thought: then why does A not agree to provide P with the information on his choice of $\underline{a}$? The answer is of course that we then have a problem of incentive compatibility: if the contract is made conditional on A's <u>report</u> of $\underline{a}$ then A has an incentive to manipulate this information to his own advantage.

[27]For example, suppose A can choose $\underline{a}$ from the set $\{1,2,3\}$, and the following distributions of x and z are then possible:

$a =$	1	2	3
$x = 0$	.8	.5	.2
$x = 1$	.2	.5	.8
$z = 0$	.9	.5	.1
$z = 1$	.1	.5	.9

where x can be only either 0 or 1 and likewise for z. Then under a contract based on x alone, A could be rewarded for a "high" effort level, even when he sets $\underline{a} = 1$, with a probability of .2, but this could only happen with a

probability of .02 if z is incorporated into the contract. Likewise A could be penalized for a low value of x with a probability of .2 even if he had set $\underline{a} = 3$, while this probability falls to .02 if z is incorporated. Thus, use of the information on z allows better design of the contract.

References

Borch, K., "Equilibrium in a Reinsurance Market," Econometrica, Vol. 30, No. 3 (1962), pp. 424-444.

Davis, O. A. and A. B. Whinston, "Welfare Economics and the Theory of the Second Best," Review of Economic Studies, (1962).

Grossman, S. J. and O. D. Hart, "An Analysis of the Principal-Agent Problem," Econometrica, Vol. 51, No. 1 (1983), pp. 7-45.

Hammond, P., "Straightforward Individual Incentive Compatibility in Large Economies," Review of Economic Studies, Vol. 46 (1979), pp. 263-282.

Harris, M. and A. Raviv, "Optimal Incentive Contracts with Imperfect Information," Carnegie Mellon University, mimeo (1976).

Harris M. and A. Raviv, "Optimal Incentive Contracts with Imperfect Information," Journal of Economic Theory, Vol. 20 (1979), pp. 231-259.

Holmström, B., "Moral Hazard and Observability," Bell Journal of Economics, Vol. 10 (1979), pp. 74-91.

Lipsey, R. G. and K. Lancaster, "The General Theory of the Second Best," Review of Economic Studies (1956/7).

Mirrlees, J., "Notes on Welfare Economics, Information and Uncertainty," in Balch, McFadden and Wu (eds.), Essays in Economic Behavior Under Uncertainty, Amsterdam, North Holland Publishing Co. (1974).

Mirrless J. A., "The Theory of Moral Hazard and Unobservable Behavior--Part I," Nuffield College, Oxford, mimeo (1975).

Ross, S., "The Economic Theory of Agency: The Principal's Problem," American Economic Review, Vol. 63 (1973), pp. 134-139.

Ross, S., "On the Economic Theory of Agency and the Principle of Similarity," in Essays in Economic Behavior Under Uncertainty, Balch, McFadden and Wu (eds.), North Holland Publishing Co. (1974).

Shavell, S., "Risk-sharing and Incentives in the Principal-Agent Relationship,." Bell Journal of Economics, Vol. 10 (1979), pp. 55-73.

Spence, M. and R. Zeckhauser, "Insurance, Information and Individual Action," American Economic Review, Vol. 61 (1971), 380-387.

Figure 1 Optimal Risk-Sharing

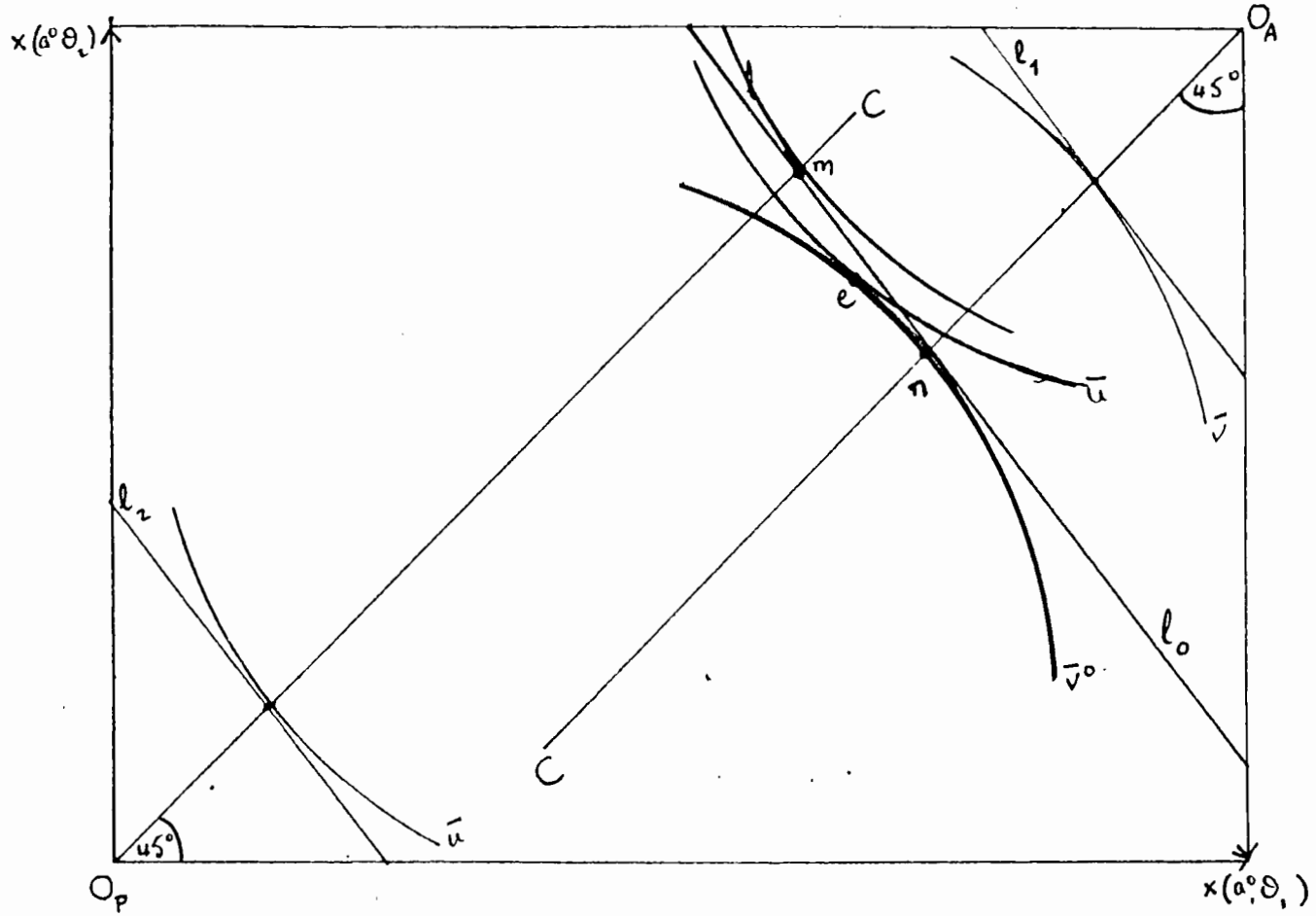

Figure 2. Optimal a with Risk-neutral.

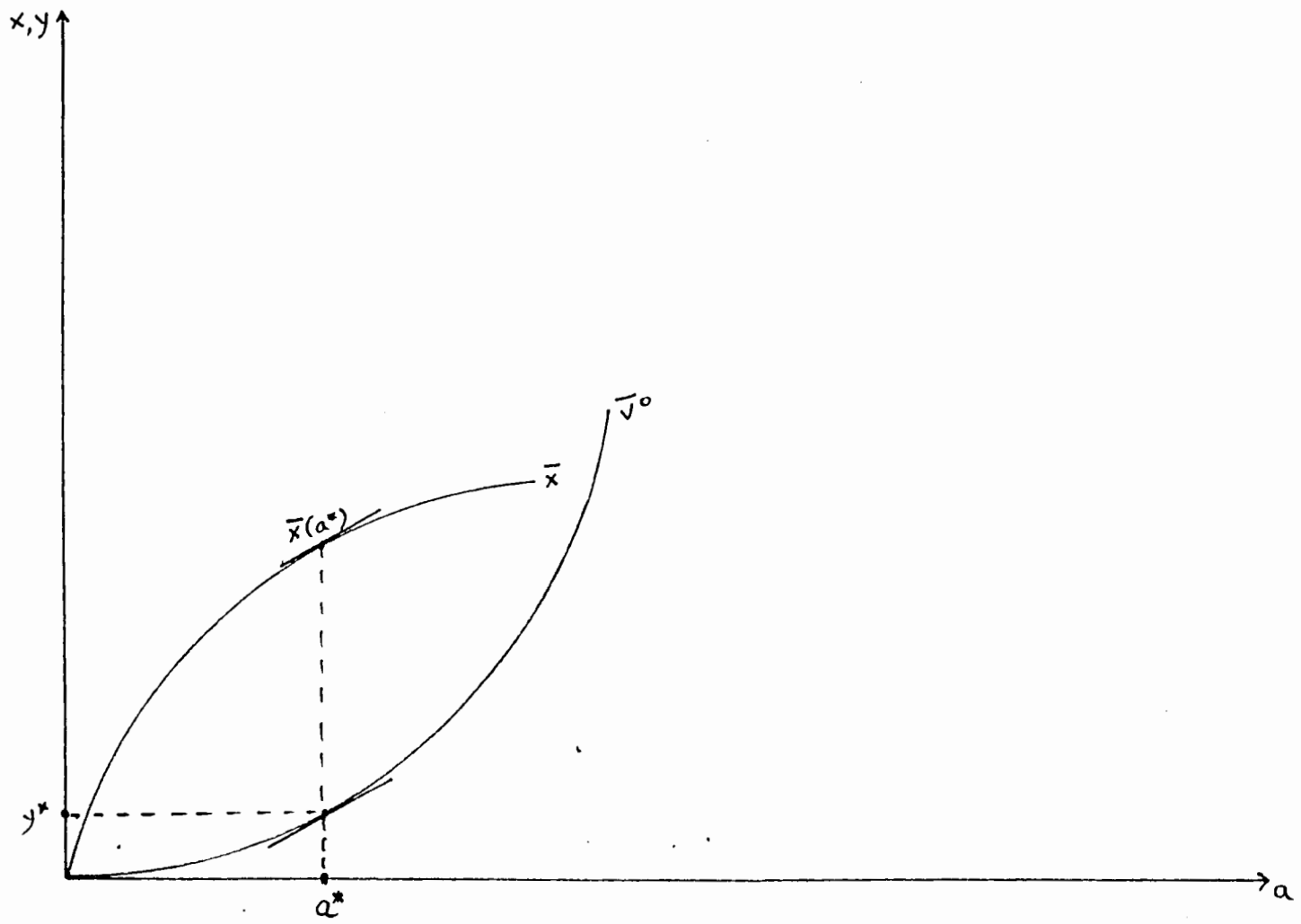

Figure 3. Observable $\alpha = a + \varepsilon$

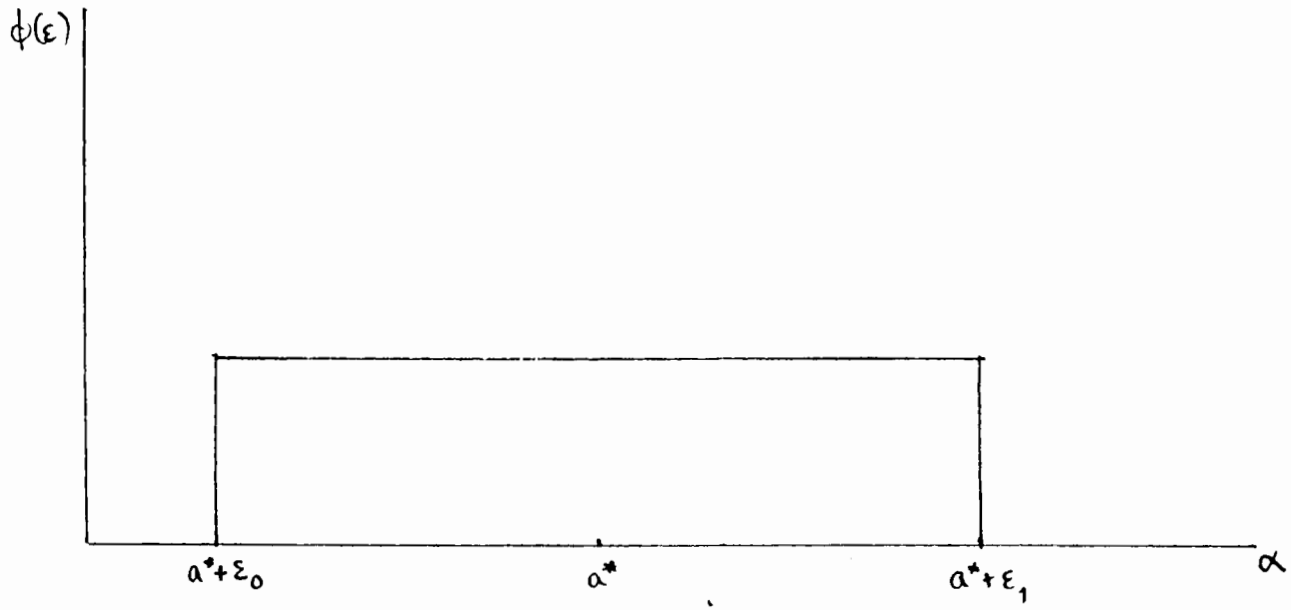

Figure 4 : Non-uniqueness of a for given η.

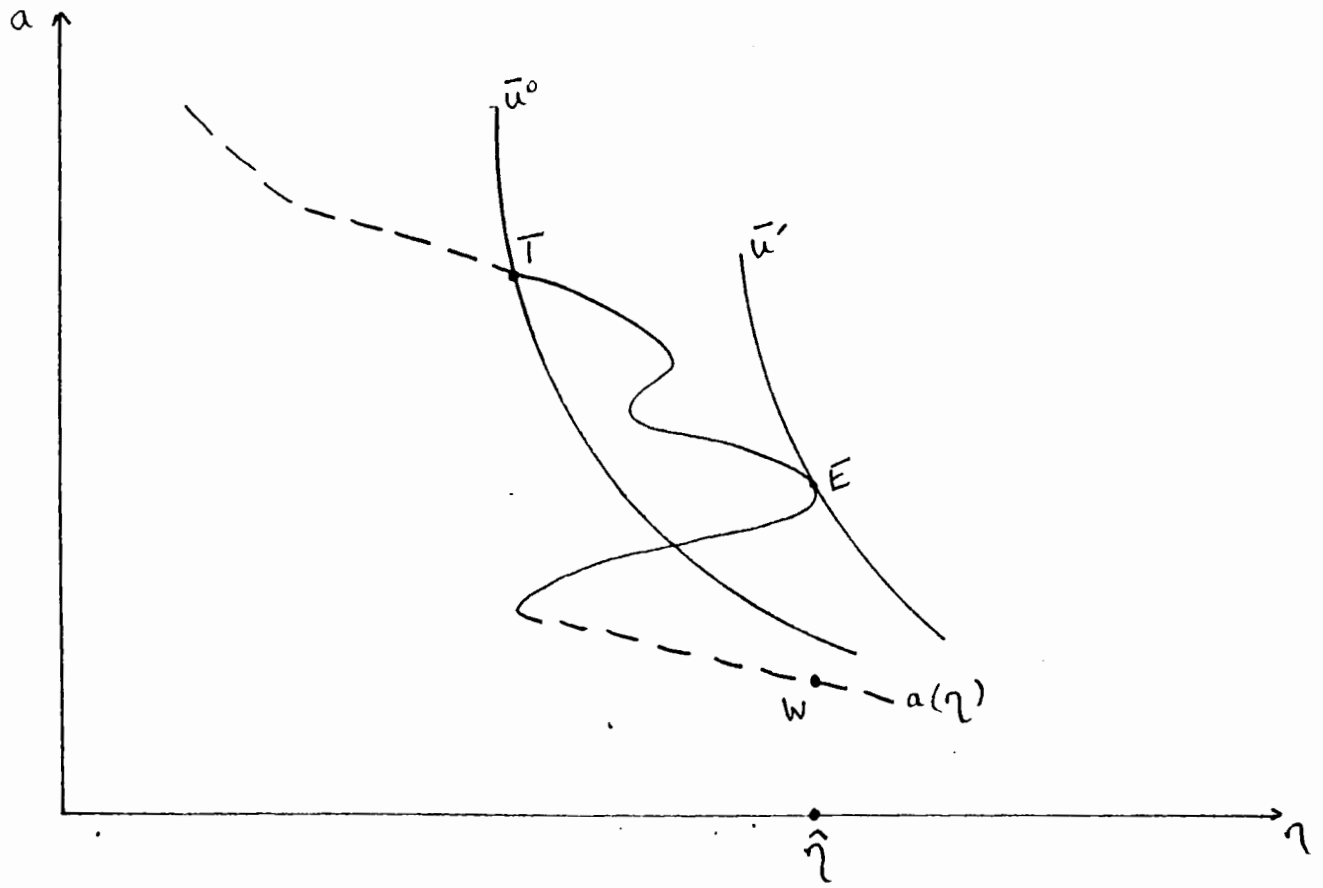

Figure 5. Change in $\phi(x,a)$ as a increases