Risk sharing and incentives in the principal and agent relationship

Steven Shavell
Department of Economics
Harvard University

This article studies arrangements concerning the payment of a fee by a principal to his agent. For such an arrangement, or fee schedule, to be Pareto optimal, it must implicitly serve to allocate the risk attaching to the outcome of the agent's activity in a satisfactory way and to create appropriate incentives for the agent in his activity. Pareto-optimal fee schedules are described in two cases: when the principal has knowledge only of the outcome of the agent's activity and when he has as well (possibly imperfect) information about the agent's activity. In each case, characteristics of Pareto-optimal fee schedules are related to the attitudes toward risk of the principal and of the agent.

1. Introduction

■ Many economic arrangements which involve problems of risk sharing and incentives may be described in terms of the principal and agent relationship. As previous writers have observed, examples include not only the relationship between a professional and his client but also that between insurer and insured, shareholders and management, and even society and a polluting firm.[1] In all these cases, one party, the principal, "enjoys" the outcome of the activity of the other party, the agent. The agent's effort (or expenditure or, more generally, his action) together with a random element determines the outcome. The principal then pays the agent a fee. For the case of a professional and his client this description of the principal and agent relationship is obviously appropriate. The description may be seen to apply to the other cases also and, indeed, to any relationship where only one of the parties directly influences the probability distribution of the outcome.[2]

Our interest here is in seeing how the fee would be expected to relate to the outcome and, if the principal has information about the agent's effort, to that

---

I wish to thank K. Arrow, P. Diamond, A.M. Polinsky, R. Radner, L. Weiss, and M. Weitzman for comments and the National Science Foundation (grant no. SOC76-20862) for financial support.

[1] See Ross (1973), Spence and Zeckhauser (1971), Stiglitz (1974, 1975), Mirrlees (1976), and Harris and Raviv (1976, 1978a, b). See also Shavell (forthcoming), which proves for a finite state model of moral hazard and insurance many of the results of this paper.

[2] This becomes clear if we allow ourselves to think of the outcome or the fee as being negative or positive. For example, consider the case of society (the principal) and a polluting firm (the agent). In this case the outcome is negative—it is the loss to society caused by an accident resulting in pollution—and the fee paid by society is also negative—it is the fine paid by the firm.

as well. Consider first the situation when the principal has no information about effort, so that the fee can depend only on the outcome. Then it is well known that (i) if the agent is risk neutral, his fee would equal the outcome minus a constant, the principal's share. Such a fee allocates risk in a desirable way (the principal might be risk averse) and provides the right incentive to the agent. But it is also well understood that if the agent is risk averse, such a fee would have the disadvantage of subjecting him to the risk associated with the outcome, given his level of effort. On the other hand, insuring the agent against this risk by setting the fee equal to a constant would leave him with no incentive to take effort. We prove that (ii) if the agent is risk averse, his fee would always depend to some extent on the outcome, but it would never leave him bearing all the risk. We also show that (iii) the achievable level of welfare approaches the first-best level as the "efficiency" of the agent's effort either tends to zero or grows large—in other words, problems owing to inappropriate incentives disappear. These results (and the following ones) hold whether the principal is risk neutral or risk averse.

Assume now that the principal has information about the agent's effort, so that the fee may depend on both the outcome and the indicator of effort. In this case, it is known that (iv) if the agent is risk neutral, nothing is lost if his fee depends on the outcome alone—in the way described in (i). Therefore, information about effort has no value.[3] But if the agent is risk averse, there would be a clear advantage to giving him an incentive by making his fee depend at least in part on effort—provided that it is observed with complete accuracy—rather than solely on the risky outcome. In fact, if effort is observed with complete accuracy, a first-best solution is attainable: make the fee very low unless the agent chooses exactly the first-best level of effort, and given that this is done, have the fee allocate the outcome in a way which takes full advantage of the possibilities for mutually beneficial risk sharing. However, we are not assuming that effort is necessarily observed with complete accuracy. Thus, the use of information about effort would introduce the new risk that the fee might reflect an inaccurate perception of the agent's true effort. Other things equal, the introduction of a new risk is, of course, undesirable for the agent; and, if the principal is risk averse, it is undesirable for him too. Consequently, there is a real question whether the information is useful. This question was initially posed by Harris and Raviv (1976, 1978a), who concluded that only under certain very limited conditions would the information be useful.[4] However, we prove that (v) if the agent is risk averse, his fee would always depend to some extent on information which the principal has about his effort—the information is always of value;

---

[3] Arrow (1971a) and Stiglitz (1974) appear to take this for granted. Harris and Raviv (1976) give the first proof of which I am aware.

[4] In the Harris-Raviv model the principal observes effort e plus a random variable η with zero mean and with a positive density on an interval $[-a,b]$ and a zero density elsewhere. As a result, it turns out that not only is information necessarily useful, but it actually allows a first-best solution: Let e^* be the first-best e and make the fee very low if $e + \eta < e^* - a$. Otherwise, make the fee depend only on the outcome and set it so that the risk is appropriately shared. Then the agent will choose $e = e^*$, for he will not dare choose $e < e^*$, and if effort involves disutility, he will not choose $e > e^*$ either.

However, Harris and Raviv (1978b) revised their 1976 paper. This later paper acknowledges that information is always valuable and emphasizes the possible usefulness of "dichotomous" fee arrangements, those which penalize the agent (in that the amount received falls discontinuously) when the observed level of effort lies in an unacceptable region.

this result has been independently proved by Holmström (1979). But note that it follows from (iii) above that the value of information approaches zero as the efficiency of effort becomes small or grows large.

After proving these results, we conclude by discussing informally several examples of principal and agent relationships.

2. The problem of the principal and the agent

■ Define the following notation:

$U(\cdot)$ = principal's utility function (in wealth)
$V(\cdot,\cdot)$ = agent's utility function (in wealth and effort)
e = agent's effort
z = principal's observation of e
x = outcome
$\phi(\cdot)$ = fee paid by the principal to the agent (a function of x alone or of x and z, as specified below)
$r(x; e)$ = probability density of x given e
$q(z|x; e)$ = probability density of z given x and e.

The principal and agent are each assumed to act so as to maximize expected utility. While the principal's utility is assumed to depend only on wealth, the agent's may depend on "effort" as well. We may interpret effort as nonmonetary, as nonmonetary but having a monetary equivalent, or as an expenditure; in the latter two cases, if wealth is w, let $V(w,e) = V(w - e)$. It is assumed that V and U are continuously differentiable and are increasing and linear or strictly concave in wealth and also that V is decreasing in effort, a nonnegative variable.[5]

It is also assumed that the outcome, which is monetary, depends on effort and a random "state of nature" and that a higher level of effort results in a higher outcome no matter what the state of nature. In addition, it is assumed that the agent makes his decision before the state of nature becomes known.

The principal may or may not have information about the agent's effort. If he does, it is given by z, which depends on effort and the state of nature.[6] (The principal and the agent would often have an opportunity to improve the quality of z at some cost but we do not study this problem explicitly.) Therefore, the distribution of z might depend on x as well as on e.[7] It is assumed that z conveys information about e in the sense that conditional on any outcome, *different levels of true effort are associated with different probability density functions q of observed effort.*[8] In fact, it will be assumed that the dependence

---

[5] The assumption that effort involves disutility does not make this model different from that of Ross (1973) in any important way. (In Ross' model, the agent merely chooses a random variable from a set of alternatives—there is no effort expended in the choice.) Most of our results would carry over to his model; see footnote 15 below.

[6] Thus, given e, the distribution of the state of nature θ induces a joint distribution over x and z through $x = x(e,\theta)$ and $z = z(e,\theta)$.

[7] Consider, for example, the problem of moral hazard in the context of fire insurance. The difficulty faced by an insurer (the principal) in determining what precautions (e) an insured (the agent) took to prevent a fire might depend on its severity (x), since it might have destroyed valuable evidence.

[8] Observe that if this were not true, then knowledge of z would not alter a posterior (based on x and a prior over e) subjective probability distribution over e. In this footnote, let d stand for the relevant probability density. We are to show that if $d(z|x,e) = d(z|x)$, then $d(e|x,z) = d(e|x)$.

on e of the density is sufficiently smooth so that all the moments of z (conditional on x and e) are differentiable in e. Now the moments of a probability density uniquely identify the density.[9] Thus, some moment of z must change with e, and it will be assumed that the derivative with respect to e of such a moment is nonzero.

The fee can depend only on variables known to both parties. Since it is natural to assume that the agent knows whatever the principal knows, the fee is assumed to be a function of the variables known by the principal.

Given a fee schedule, the agent selects effort to maximize expected utility. Thus, if the principal knows only x, so that $\phi = \phi(x)$, the agent maximizes over e

$$EV(\phi,e) = \int V(\phi(x),e)r(x;e)dx. \tag{1}$$

If the principal knows x and z, the agent maximizes over e

$$EV(\phi,e) = \iint V(\phi(x,z),e)q(z|x;e)dz\, r(x;e)dx. \tag{2}$$

We shall assume that there is a unique solution to the agent's problem.

Given a fee schedule and the agent's effort, the principal's expected utility is

$$EU(\phi,e) = \int U(x - \phi(x))r(x;e)dx \tag{3}$$

if $\phi = \phi(x)$; and $EU(\phi,e)$ is defined analogously in the case where $\phi = \phi(x,z)$.

In other words, given a fee schedule, the expected utilities of the agent and the principal are determined. We would expect that the fee schedule actually employed by them would be such that no alternative schedule would be mutually advantageous.[10] Thus, the problem of the principal and the agent as considered here is to find a *Pareto optimal fee schedule*. Formally, ϕ is Pareto optimal if it solves the problem[11]

$$\max EU(\phi,e) \text{ over } \phi \tag{4}$$

subject to

$$EV(\phi,e) \geqq V^0 \tag{5}$$

$$EV(\phi,e) \text{ is maximized over } e. \tag{6}$$

Now $d(e|x,z) = d(e,x,z)/d(x,z) = d(z|x,e)d(x,e)/d(x,z) = d(z|x)d(x,e)/d(x,z) = d(z|x)d(x,e)/(d(z|x)d(x)) = d(x,e)/d(x) = d(e|x)$.

Note also that if $x(e,\theta)$ is strictly monotonic in θ as well as in e, then if $d(z|x,e)$ depends on e, $d(z|x,\theta)$ must depend on θ and conversely. That is, "*z conveys information about e*" *is equivalent to* "*z conveys information about θ.*" This follows immediately from the fact that if $x(e_1,\theta_1) = x(e_2,\theta_2)$, then $e_1 \neq e_2 \Leftrightarrow \theta_1 \neq \theta_2$.

[9] This is true under quite general conditions, for example, when the range of the random variable is finite. See Rao (1965, p. 86).

[10] This presumes that the parties do not engage in mutually destructive bargaining and that they possess the information necessary to compute each other's expected utility.

[11] Of course, the problem could be stated in the equivalent form, maximize $EV(e,\phi)$ over ϕ subject to $EU(e,\phi) \geqq U^0$ and $EV(e,\phi)$ is maximized over e.

We assume that the problem has a solution. However, Mirrlees (1974) has analyzed a similar problem for which there was no solution. A solution failed to exist for his problem because he assumed that utility was unbounded (as it happened, from below). There are well-known reasons for taking utility to be bounded if expected utility is to be used as an indicator of preference over uncertain prospects. (For example, if utility is unbounded, one can easily construct prospects

The constant V^0 in (5) is determined by bargaining power or by market forces. In the latter case, it is the utility level which the agent could achieve by going elsewhere. It is assumed that the first-order condition

$$EV_e(\phi,e) = 0 \tag{7}$$

identifies the solution to (6).[12]

The problem of maximizing $EU(\phi,e)$ over ϕ and e subject only to (5) will be referred to as the first-best problem, since its solution describes a Pareto optimum when e as well as ϕ can be directly chosen.

3. The agent's fee

■ **The principal knows only the outcome: the fee $\phi = \phi(x)$.** The following result seems to be well known and is intuitively clear; we shall defer a proof of it and discussion until we present Proposition 4.

Proposition 1. Suppose that the agent is risk neutral. Then (corresponding to any V^0) there is a Pareto optimal fee schedule under which the agent is paid the outcome minus a constant (specifically, $\phi(x) = x - k$, where k is the principal's share).

Let us now consider the case of a risk averse agent.

Proposition 2. Suppose that the agent is risk averse. Then under a Pareto optimal fee schedule the agent (a) is paid an amount which must depend to some extent on the outcome, but (b) he never bears all the risk.

The idea behind the proof of (a) is straightforward. Assume first that the principal is risk neutral and suppose that the fee equals a constant. Now alter the fee schedule by making it depend very slightly on the outcome. *There will be no first-order effect on the agent's expected utility which can be attributed to the imposition of risk* because, initially, his fee, and thus his wealth, was constant.[13] However, if (7) holds[14] and the fee schedule is altered in the appropriate way, *there will be a positive first-order effect on the agent's effort.* By the envelope theorem, this change in effort will have no first-order effect on the

---

A and B such that A is strictly better than B no matter what the state of nature, yet A and B cannot be distinguished on the basis of expected utility.) See Savage (1972, pp. 81–82) and Arrow (1971b, pp. 63–69).

Mirrlees showed directly that a solution for his problem failed to exist by demonstrating that the first-best solution could be arbitrarily closely approximated. Under his construction (and in his notation), individuals are punished if their output y falls below a level η. By raising effort z, they can lower the chance of a $y \leqq \eta$. No matter how small η is, there exists a potential punishment sufficiently large to induce individuals to choose an appropriately high z, say z^*. And because η may be made very small, the probability of punishment may be made very low. Finally, it turns out that the loss in expected utility owing to the possibility of punishment goes to zero as η gets small, even though the size of punishment (in utils) grows infinite. In other words, an appropriate incentive is achieved through the use of a scheme which involves an arbitrarily low "cost" in terms of expected utility. If utility is bounded, it is obvious that this construction is not possible because the incentive effect of a bounded punishment goes to zero as the probability of punishment goes to zero.

[12] See Mirrlees (1975) for a discussion of difficulties which may arise when there may be multiple solutions to the first-order condition.

[13] Let W be the utility function of a risk averse individual and let y be his initial wealth. Suppose that a scalar multiple t of a random variable η which has zero mean is added to y. Then expected utility as a function of t is $EW(t) = \int W(y + t\eta)dG(\eta)$ and $EW'(t) = \int \eta W'(y + t\eta)dG(\eta)$, so that $EW'(0) = W'(y) \int \eta dG(\eta) = 0$. (Here G is the c.d.f. of η.) In other words, *the first-order*

agent's expected utility, but it will increase the outcome in every state of nature, allowing the principal to reward the agent and make himself better off as well. If the principal is risk averse, the argument is more complicated. It now must take into account the fact that making the fee schedule depend on the outcome may have a nonzero first-order effect on the principal's expected utility owing to the imposition of risk, since, initially, his wealth was stochastic (it equalled the outcome minus a constant).

The idea behind (b) is similar. Assume that the agent bears all the risk—that the principal gets a constant. Now give the principal a small fraction of the outcome. There will be no first-order effect on his expected utility owing to the imposition of risk, since initially his wealth was constant. However, there will be a positive first-order effect on the agent's expected utility owing to a reduction in risk-bearing. By the envelope theorem, the induced change in the agent's effort will have no first-order effect on his expected utility; nor will it have a nonzero first-order effect on the principal's expected utility since, initially, his share, being a constant, was independent of the level of effort.

Proof: To prove (a) suppose that the fee schedule equals a constant k. We shall show that a constant fee cannot be Pareto optimal by constructing a new fee schedule which makes both the principal and the risk averse agent better off. If the agent is paid a constant k, he will choose the minimum level of effort, zero. Define the new fee schedule by

$$\phi(x) = k + \alpha(x - \bar{x}(0)) + \alpha\beta, \tag{8}$$

where $\alpha > 0$ and $\beta > 0$ are to be determined and where $\bar{x}(e) = \int xr(x; e)dx$. Notice that under ϕ the agent is paid more on average if he continues to choose $e = 0$, for he always gets an additional $\alpha\beta$ and the mean of $\alpha(x - \bar{x}(0))$ is zero. On the other hand, ϕ appears to give an incentive to raise e, for if $e > 0$, $\bar{x}(e)$ will exceed $\bar{x}(0)$.

The agent now maximizes over e,

$$j(e,\alpha,\beta) = \int V(k + \alpha(x - \bar{x}(0)) + \alpha\beta, e)r(x; e)dx \tag{9}$$

and the first-order condition,

$$j_e(e,\alpha,\beta) = 0, \tag{10}$$

determines the optimal e as a function of α and β; e will occasionally be written as $e(\alpha,\beta)$. Let $J(\alpha,\beta) = \max_e j(e,\alpha,\beta)$, so that

$$\begin{aligned} J_\alpha(\alpha,\beta) &= j_e(e,\alpha,\beta)e_\alpha(\alpha,\beta) + j_\alpha(e,\alpha,\beta) \\ &= j_\alpha(e,\alpha,\beta) = \int (x - \bar{x}(0) + \beta)V_1(k + \alpha(x - \bar{x}(0)) + \alpha\beta, e)r(x; e)dx. \end{aligned} \tag{11}$$

---

effect of the imposition of risk is zero. This is as expected, since a differentiable function is by definition linear (and therefore displays risk neutrality) in the small.

[14] If the agent is paid a constant k, he will choose the minimum level of effort, zero. Now $V_e(k,0) = 0$ might hold, as would be the case if e is, say, attention and the marginal disutility of increasing attention above some natural zero level is zero. But if e is an expenditure, $V(k,e) = V(k - e)$, so $V_e(k,0) = -V'(k) < 0$; thus, the assumption that (7) holds is not an innocuous one with regard to part (a).

Thus, since $e(0,\beta) = 0$,

$$J_\alpha(0,\beta) = j_\alpha(0,0,\beta) = \int (x - \bar{x}(0) + \beta) V_1(k,0) r(x;\, 0)dx = \beta V_1(k,0) > 0. \quad (12)$$

Consequently, for any positive β, if α is chosen small enough, the agent will be better off.

To complete the proof of (a) we must show that the principal as well can be made better off. His expected utility is

$$K(\alpha,\beta) = \int U(x - k - \alpha(x - \bar{x}(0)) - \alpha\beta) r(x;\, e)dx, \quad (13)$$

where $e = e(\alpha,\beta)$. Therefore, we need to show that $K_\alpha(0,\beta) > 0$, if β is appropriately chosen. Now

$$K_\alpha(0,\beta) = e_\alpha(0,\beta) \int U(x - k) r_e(x;\, 0)dx$$
$$+ \int (\bar{x}(0) - x) U'(x - k) r(x;\, 0)dx - \beta \int U'(x - k) r(x;\, 0)dx. \quad (14)$$

However,

$$\int (\bar{x}(0) - x) U'(x - k) r(x;\, 0)dx$$
$$= \int_{x \leqq \bar{x}(0)} (\bar{x}(0) - x) U'(x - k) r(x;\, 0)dx + \int_{x > \bar{x}(0)} (\bar{x}(0) - x) U'(x - k) r(x;\, 0)dx$$
$$\geqq \int_{x \leqq \bar{x}(0)} (\bar{x}(0) - x) U'(\bar{x}(0) - k) r(x;\, 0)dx$$
$$+ \int_{x > \bar{x}(0)} (\bar{x}(0) - x) U'(\bar{x}(0) - k) r(x;\, 0)dx$$
$$= U'(\bar{x}(0) - k) \int (\bar{x}(0) - x) r(x;\, 0)dx = 0. \quad (15)$$

Also $\int U(x - k) r_e(x;\, 0)dx > 0$, since by assumption an increase in e raises x in every state of nature.

Thus, we need only show that $e_\alpha(0,\beta)$ is greater than some $\delta > 0$ no matter how small β becomes. Differentiating (10), we get

$$e_\alpha(0,\beta) = \frac{-j_{e\alpha}(0,0,\beta)}{j_{ee}(0,0,\beta)}, \quad (16)$$

and since $j_{ee}(0,0,\beta) < 0$ (the second-order condition for a regular maximum) and is independent of β, it is enough to show that $j_{e\alpha}(0,0,\beta)$ is greater than some $\delta > 0$, no matter how small β is. As

$$j_e(e,\alpha,\beta) = \int V(k + \alpha(x - \bar{x}(0)) + \alpha\beta,\, e) r_e(x;\, e)dx$$
$$+ \int V_2(k + \ldots, e) r(x;\, e)dx, \quad (17)$$

we have

$$j_{e\alpha}(e,\alpha,\beta) = \int (x - \bar{x}(0) + \beta)V_1(k + \alpha(x - \bar{x}(0)) + \alpha\beta, e)r_e(x; e)dx$$
$$+ \int (x - \bar{x}(0) + \beta)V_{21}(k + \ldots, e)r(x; e)dx. \quad (18)$$

As $\int r(x; e)dx = 1$, we have $\int r_e(x; e)dx = 0$. Using this, we have

$$j_{e\alpha}(0,0,\beta) = V_1(k,0)\int xr_e(x; 0)dx + \beta V_{21}(k,0). \quad (19)$$

But, since x is increasing in e in every state of nature, we have $\int xr_e(x; 0)dx > 0$. Thus $j_{e\alpha}(0,0,\beta)$ is of the form $K_1 + \beta K_2$ with $K_1 > 0$. Consequently, let $\delta = K_1/2$, so that $j_{e\alpha}(0,0,\beta) > \delta$ for all small β.

To prove (b), suppose that the principal receives a constant k, so that the agent gets $x - k$, and denote by $\hat{e}$ the level of effort taken by the agent in this case. We shall construct a new fee schedule that makes both the risk averse agent and the principal better off. Define the new schedule by

$$\phi(x) = (1 - \alpha)x - k + \alpha\beta, \quad (20)$$

where $\alpha > 0$ and $\beta > 0$ are to be determined. The agent selects e to maximize

$$j(e,\alpha,\beta) = \int V((1 - \alpha)x - k + \alpha\beta, e)r(x; e)dx. \quad (21)$$

If $J(\alpha,\beta) = \max_e j(e,\alpha,\beta)$, then

$$J_\alpha(0,\beta) = \int (\beta - x)V_1(x - k, \hat{e})r(x; \hat{e})dx, \quad (22)$$

which is greater than zero if

$$\beta > \frac{\int xV_1(x - k, \hat{e})r(x; \hat{e})dx}{\int V_1(x - k, \hat{e})r(x; \hat{e})dx}. \quad (23)$$

Now we wish to show that with such a β the principal also can be made better off if α is chosen small enough. His expected utility is

$$K(\alpha,\beta) = \int U(k + \alpha x - \alpha\beta)r(x; e)dx, \quad (24)$$

where e is chosen by the agent. Now

$$K_\alpha(0,\beta) = U'(k)(\bar{x}(\hat{e}) - \beta). \quad (25)$$

Thus,

$$\beta < \bar{x}(\hat{e}) \quad (26)$$

must hold for the principal to be made better off. A β satisfying (23) and (26) can be found if

$$\frac{\int xV_1(x - k, \hat{e})r(x; \hat{e})dx}{\int V_1(x - k, \hat{e})r(x; \hat{e})dx} < \bar{x}(\hat{e}) \quad (27)$$

holds, but this may be proved using steps analogous to those in (15). *Q.E.D.*

It should be noted that Proposition 2 is consistent with the characteristics one would expect of a fee schedule that shared risk in a Pareto optimal way between a risk averse principal and a risk averse agent. As Borch (1962) originally pointed out, a necessary condition for Pareto optimal risk sharing is that $V_1(\phi(x),e)/U'(x - \phi(x))$ is constant. If $V_{11} < 0$ and $U'' < 0$, this condition rules out the possibility that either $\phi(x)$ or $x - \phi(x)$ is constant. However, it is shown in the proof of Proposition 5 that a Pareto optimal fee schedule depending only on the outcome cannot share risk in a Pareto optimal way.

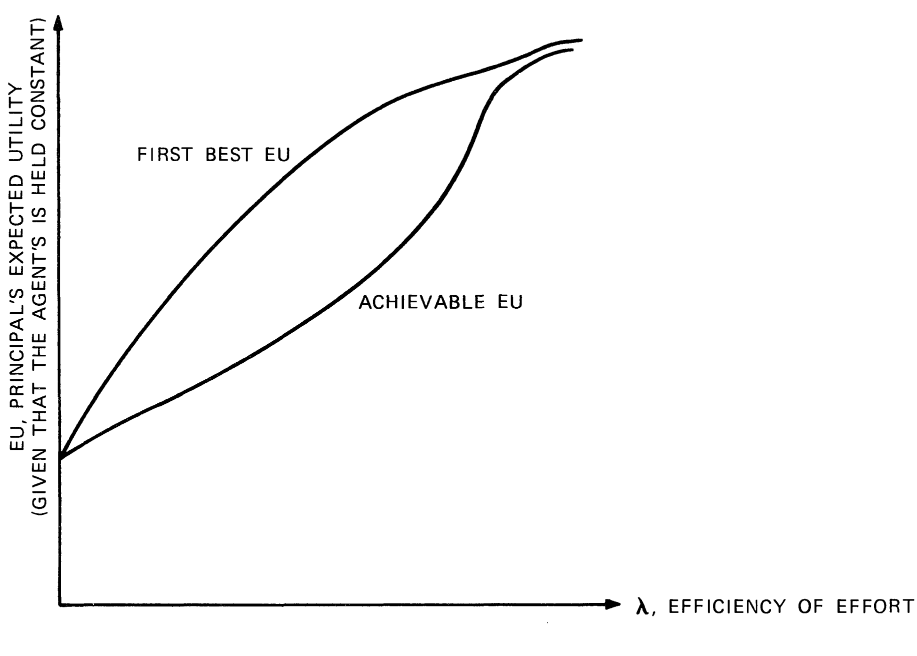
The next proposition concerns the relation between an index of the efficiency of the agent's effort and the solution to the problem of the principal and the agent. We proceed as follows. If the index is λ and the level of effort is e, the density of x will be assumed to be $r(x; \lambda e)$. We assume that as $\lambda e \rightarrow \infty$ the density function converges uniformly to a density denoted r^*. Our definition of efficiency has a simple motivation. Suppose that r is determined by the quantity of some good purchased by the agent and that e is expenditure on the good. Then if p is the price of the good, the quantity purchased is e/p, so the density is $r(x; e/p)$ and $\lambda = 1/p$.

If $\lambda = 0$, r is not affected by e and the problem is merely one of risk sharing, so a first-best solution is achievable. When $\lambda > 0$, r is affected by e and a first-best solution is not achievable (for, as noted above, risk is not shared in a Pareto optimal way). However, Proposition 3 shows that as $\lambda \rightarrow \infty$, the difference between the achievable solution and the first-best solution tends to zero. This is illustrated in Figure 1.

The intuition behind the result is that when λ is high, little effort is required to change radically the density, to bring it near its best level r^*. Thus, at most, only a slight deviation of the fee schedule from a first-best schedule is needed to overcome a problem of an incorrect incentive to take effort.

FIGURE 1

FIRST BEST AND ACHIEVABLE SOLUTIONS



Proposition 3: The difference in welfare between the first-best and the achievable solutions to the problem of the principal and the agent tends to zero as the efficiency of effort grows large.

Specifically, if $EU^*(\lambda)$ and $EU(\lambda)$ denote the expected utility of the principal under the first-best and under the achievable solutions, respectively, then

$$\lim_{\lambda \to \infty} (EU^*(\lambda) - EU(\lambda)) = 0.$$

The proof is presented in the Appendix.

□ **The principal knows the outcome and has information about effort: the fee $\phi = \phi(x,z)$.** As noted above, the following result, which we state and prove for the sake of completeness, has been taken for granted by several writers and was first proved by Harris and Raviv (1976). Our proof of it (and of Proposition 1) is almost an immediate consequence of Jensen's inequality and is presented in the Appendix.

Proposition 4: Suppose that the agent is risk neutral. Then (corresponding to any V^0) there is a Pareto optimal fee schedule under which the agent is paid an amount that depends only on the outcome (specifically, the schedule is of the form $\phi(x,z) = x - k$ where k is a constant amount kept by the principal). Thus information about the agent's effort is of no value.

This result is intuitively clear. Because the agent is risk neutral, he can act in behalf of the principal, who may be risk averse, as a perfect insurer against the risk of variation in the "net return" (the outcome minus the "cost" of effort). Thus, from the point of view of both parties, it is desirable for the agent to maximize the expected net return. If the agent's fee is equal to the outcome minus a constant, the principal's share, he will in fact maximize the expected net return since he bears the cost of his effort and is risk neutral.

There is only one reason why it might have been valuable to make the fee depend on information about the agent's effort: to reduce the variability of his fee, given his effort. But, again, because he is risk neutral, this concern is not relevant.

Let us now consider the alternative case.

Proposition 5: Suppose that the agent is risk averse. Then under a Pareto optimal fee schedule the agent is paid an amount which must depend to some extent on information about his effort; thus the information is of positive value.

Because the agent is risk averse, it is desirable for him to avoid risk whenever possible. In particular, he would be better off—and, as explained in the introduction, the first-best solution would be attainable—if his actual level of effort, rather than the risky outcome, served as an incentive in the fee schedule. On the other hand, if the principal has only imperfect information about effort and uses that information, a new risk is introduced. To prove that the fee schedule should depend on the information despite the new risk, an argument similar to that used in Proposition 2 is employed. Suppose that the fee schedule depends solely on the outcome. Then alter the schedule by making it depend very slightly on the imperfect information about the agent's behavior. There will be no first-order effect on the expected utility of the agent or principal which can be attributed to the imposition of a new risk, since, initially, the wealth of each was constant *given* the outcome. By the envelope theorem any change in effort will have no first-order effect on the expected utility of the agent. However, if the schedule is altered in the appropriate way, the change in effort will have a positive first-order effect on the expected utility of the principal and some of his benefits can be given to the agent to make him better off as well. The proof is presented in the Appendix.[15]

Although Proposition 5 shows that information about effort is valuable if the agent is risk averse, it does not tell us how valuable. However, it follows from Figure 1 that *the value of information goes to zero as the efficiency of effort goes to zero or to infinity.*

4. Examples

■ To conclude, let us briefly discuss four examples involving principal and agent relationships.

□ **Strict liability vs. negligence standards in the control of stochastic externalities.** Consider a situation in which firms occasionally cause accidents having adverse external effects (environmental damage is an outstanding example) and in which the probability or severity of the accidents is influenced to some degree by the actions of the firms. In such a situation, suppose the government wishes to intervene, employing either a strict liability approach or negligence standards. Under strict liability, the government charges firms for the losses imposed on society whenever accidents occur. Under negligence standards, the government charges firms for the losses only when accidents involve negligence—failure to meet the standards.

The choice between strict liability and negligence may be thought of as a choice of fee schedule for a principal (the government or society) and an agent (a firm). In this view, under strict liability the agent's fee (zero if there is no accident and negative otherwise) depends only on the outcome (accident losses), whereas under a negligence standard the fee depends on information about the agent's actions. Suppose that it is appropriate to regard the management of a typical firm as risk neutral with regard to the losses caused by an accident. This might be true if the potential losses are small in relation to the total business of the firm. Then the results indicate that a strict liability approach would work well. However, if it is appropriate to regard the management as risk averse, the results suggest that there might be a tendency to favor negligence standards. The reason is that whereas under strict liability the firm might take excessive care to prevent accidents or even decide to exit from the industry,[16] with negligence standards the firm would not be subject to the risk of having to pay if an accident occurred, provided that the firm met the standards.

Of course, the choice between strict liability and negligence standards depends not only on the management's attitude toward risk but also on the quality of information that the government can hope to acquire about the firms' ad-

[15] Subject to one qualification, the proof given there would establish the result for Ross' model (or for our model, modified to allow for $r(x; e)$ to change with e in an arbitrary way). We would have to add the assumption that under a Pareto optimal fee schedule depending only on the outcome, risk is not shared in a Pareto optimal way. As Ross shows, such Pareto optimal risk sharing could occur in his model, but the last step of our proof rules it out for our model.

[16] If the firms can cheaply purchase insurance against the liability, the story becomes more complicated (Shavell, 1978).

herence to standards, on the cost of obtaining this information, and on a variety of other considerations.

□ **Moral hazard and insurance.** In this familiar situation, the actions of an insured individual may influence the probability distribution of loss. If the insurer does not have information about the insured's actions, then our results imply that although the resulting problem of moral hazard cannot itself eliminate the market for insurance (Proposition 2(b)), the coverage sold would not be complete (Proposition 2(a)). However, when the cost of taking actions to reduce accident risks is either very high or very low (corresponding to very low or very high λ), the welfare loss due to moral hazard is likely to be small and the level of coverage nearly full. (See Figure 1.) If the insurer has information which reflects something of the truth about the insured's actions, then our results imply that no matter how imprecise the information, it has value and ought to be incorporated into the terms of a policy.

□ **Lawyer and client.** The fee which a client pays his lawyer may be based on the time the lawyer spends on a case and on the outcome of the case. In some instances, the fee is based solely on the outcome through the exclusive use of a contingent fee arrangement. This is presumably desirable when the lawyer is risk neutral, or at least when he is much less risk averse than his client. For example, in many medical malpractice cases the contingent fee is the only fee; in these cases it does seem reasonable to believe that the typical client has a very risk averse attitude toward the amount he has at stake, whereas the lawyer has a more nearly risk neutral attitude toward the possibility of losing his fee.[17]

In other instances, the fee is virtually never based solely on the outcome. For example, a corporate client is almost always billed by a law firm for time spent on a lengthy and complicated case. This is no doubt true because the law firm has a risk averse attitude toward the possibility of not being compensated for an effort absorbing a major part of its resources.

In most instances, as one would expect, the fee is based both on a lawyer's time and, at least indirectly, on the outcome of a case. The fee may depend indirectly on the outcome because the probability of repeat business and of new business may be influenced by the lawyer's reputation. This surely reduces the need for contingent fees.

□ **Shareholders and the manager of a firm.** If the shareholders of a firm are viewed as a principal and the manager as an agent, it is clear that the firm would not generally be operated in the way that shareholders would wish. For example, imagine that shareholders want the firm to maximize expected profits. (This would be true if shareholders hold well-diversified portfolios and if the random variables describing profits under different strategies for the firm are independent of the returns on other investments in the economy.) In other words, the shareholders as a group are a risk neutral principal. Suppose that the manager is risk averse and that his actions are not easily known to the shareholders. Then under a Pareto optimal fee schedule, the manager would not conduct the firm so as to maximize expected profits.

[17] This is especially likely if the lawyer has a large practice or is a member of a firm.

Appendix

■ **Proof of proposition 3.** The proof consists of several steps. First, let us define $EV(\phi,e,\lambda) = \int V(\phi(x),e)r(x;\ \lambda e)dx$ and define $EU(\phi,e,\lambda)$ similarly.

(i) If ϕ is strictly increasing in x, then $EV(\phi,e,\lambda)$ is strictly increasing in λ: If λ rises, λe rises, so x rises in each state of nature. Since ϕ is increasing, this means that ϕ rises in each state of nature.

(ii) If ϕ is strictly increasing in x, then (a) $\lim_{\lambda\to\infty} e(\phi,\lambda) = 0$ and (b) $\lim_{\lambda\to\infty} \lambda e(\phi,\lambda) = \infty$, where $e(\phi,\lambda)$ is the effort chosen by the agent given ϕ and λ: If (a) does not hold, then there exists $\epsilon > 0$ and a sequence $\lambda_i \to \infty$ such that $e(\phi,\lambda_i) \geqq \epsilon$ for all i. But then

$$\int V\left(\phi(x), \frac{\epsilon}{2}\right)r\left(x;\ \lambda_i \frac{\epsilon}{2}\right)dx > \int V(\phi(x),e(\phi,\lambda_i))r(x;\ \lambda_i e(\phi,\lambda_i))dx \qquad \text{(A1)}$$

for i sufficiently large. This is because

$$\begin{aligned}\lim_{i\to\infty} \int V\left(\phi(x), \frac{\epsilon}{2}\right)r\left(x;\ \lambda_i \frac{\epsilon}{2}\right)dx &= \int V\left(\phi(x), \frac{\epsilon}{2}\right)r^*(x)dx \\ &> \int V(\phi(x),\epsilon)r^*(x)dx \geqq \int V(\phi(x),e(\phi,\lambda_i))r^*(x)dx \\ &\geqq \int V(\phi(x),e(\phi,\lambda_i))r(x;\ \lambda_i e(\phi,\lambda_i))dx, \qquad \text{(A2)}\end{aligned}$$

where the last step follows from (i). But (A1) contradicts the definition of $e(\phi,\lambda_i)$, establishing (a).

If (b) does not hold, then for any $N > 0$ there exists a sequence $\lambda_i \to \infty$ such that $\lambda_i e(\phi,\lambda_i) < N$ for all i. But then

$$\int V\left(\phi(x), \frac{N}{\lambda_i}\right)r(x;\ N)dx > \int V(\phi(x),e(\phi,\lambda_i))r(x;\ \lambda_i e(\phi,\lambda_i))dx \qquad \text{(A3)}$$

for i sufficiently large. This is because

$$\begin{aligned}\lim_{i\to\infty} \int V\left(\phi(x), \frac{N}{\lambda_i}\right)r(x;\ N)dx &= \int V(\phi(x),0)r(x;\ N)dx \\ \geqq \int V(\phi(x),e(\phi,\lambda_i))r(x;\ N)dx &> \int V(\phi(x),e(\phi,\lambda_i))r(x;\ \lambda_i e(\phi,\lambda_i))dx, \qquad \text{(A4)}\end{aligned}$$

where the last step follows from (i). But (A3) contradicts the definition of $e(\phi,\lambda_i)$, establishing (b).

(iii) If ϕ allocates in a Pareto optimal way a given random variable between two risk averse parties, then ϕ must be strictly increasing. This follows immediately from Borch's equation characterizing such schedules, $A'(\phi(x))/B'(x - \phi(x)) = k$, where A and B are the utility functions and k is a constant.

(iv) Completion of proof: Let ϕ^* allocate the variable x with density r^* in a Pareto optimal way and such that $\int V(\phi^*(x),0)r^*(x)dx = V^0$. Let $EU^* = \int U(x - \phi^*(x))r^*(x)dx$. Recall that we wish to show $\lim_{\lambda\to\infty} [EU^*(\lambda) - EU(\lambda)] = 0$. Clearly, $EU^* \geqq EU^*(\lambda) \geqq EU(\lambda)$. Thus if we can establish that $\lim_{\lambda\to\infty} EU(\lambda) = EU^*$, the proof will be complete.

Assume first that the principal is risk averse. Then by (iii) ϕ^* is strictly increasing, and so is $\phi^* + k$ for any k. Thus by (ii), we have $\lim_{\lambda\to\infty} e(\phi^* + k, \lambda) = 0$ and $\lim_{\lambda\to\infty} \lambda e(\phi^* + k, \lambda) = \infty$. Hence if $k > 0$

$$\lim_{\lambda\to\infty} \int V(\phi^*(x) + k, e(\phi^* + k, \lambda))r(x; \lambda e(\phi^* + k, \lambda))dx$$

$$= \int V(\phi^*(x) + k, 0)r^*(x)dx > V^0 \quad \text{(A5)}$$

and

$$\lim_{\lambda\to\infty} \int U(x - \phi^*(x) - k)r(x; \lambda e(\phi^* + k, \lambda))dx$$

$$= \int U(x - \phi^*(x) - k)r^*(x)dx. \quad \text{(A6)}$$

To show that $\lim_{\lambda\to\infty} EU(\lambda) = EU^*$ we must show that for any $\epsilon > 0$, $EU^* - EU(\lambda) < \epsilon$ for all λ sufficiently large. Now by (A5) and (A6) it follows that if $k > 0$ is small enough, then for all λ sufficiently high $EV(\phi^* + k, e(\phi^* + k, \lambda), \lambda) > V^0$ and $EU^* - EU(\phi^* + k, e(\phi^* + k, \lambda), \lambda) < \epsilon$. But for such λ, $EU(\lambda) \geqq EU(\phi^* + k, e(\phi^* + k, \lambda), \lambda)$ since the agent's utility exceeds V^0. This implies $EU^* - EU(\lambda) < \epsilon$ for such λ.

Now assume that the principal is risk neutral. Then ϕ^* is in fact a constant k^*. Let $\bar{x} = \int xr^*(x)dx$. Now for $k_1, k_2 > 0$, $k^* + k_1(x - \bar{x}) + k_2$ is strictly increasing. And, letting this function play the role of $\phi^* + k$, an argument analogous to that of the previous paragraph completes the proof. The only step that requires comment is the analog to (A5). It is

$$\lim_{\lambda\to\infty} \int V(k^* + k_1(x - \bar{x}) + k_2, e(k^* + k_1(x - \bar{x}) + k_2, \lambda))$$

$$\times\, r(x, \lambda e(k^* + k_1(x - \bar{x}) + k_2, \lambda))dx = \int V(k^* + k_1(x - \bar{x}) + k_2, 0)r^*(x)dx$$

$$> \int V(k^*,0)r^*(x)dx = V^0. \quad \text{(A7)}$$

The last inequality is easily shown to hold for any $k_2 > 0$ and $k_1 > 0$ but sufficiently small.

□ **Proof of proposition 4.** Consider initially the solution to the first-best problem of maximizing $EU(\phi,e)$ over ϕ and e subject only to (5). Since the agent is risk neutral, we may assume that $V(w,e) = a(e)w - v(e)$, where $a(e) > 0$, so that (5) becomes

$$EV(\phi,e) = \iint a(e)\phi(x,z)q(z\,|x; e)dz\; r(x; e)dx - v(e) \geqq V^0. \quad \text{(A8)}$$

Dividing by $a(e)$, multiplying by -1, then adding $\bar{x}(e) = \int xr(x; e)dx$ to both sides and rearranging, we get

$$\iint (x - \phi(x,z))q(z\,|x; e)dz\; r(x; e)dx \leqq \bar{x}(e) - \frac{v(e)}{a(e)} - \frac{V^0}{a(e)}\,. \quad \text{(A9)}$$

Therefore, the problem is to maximize $EU(\phi,e)$ subject to (A9). But this constraint says only that the principal's mean income at most equal $\bar{x}(e) - v(e)/a(e) - V^0/a(e)$. Thus, by Jensen's inequality (see, for example, DeGroot (1970)), the maximum utility of the principal as a function of e is $U(\bar{x}(e) - v(e)/a(e) - V^0/a(e))$, which may be achieved if $x - \phi(x,z) = \bar{x}(e) - v(e)/a(e) - V^0/a(e)$. As a consequence, the optimal e, say e^*, maximizes $\bar{x}(e) - v(e)/a(e) - V^0/a(e)$. Let k^* be the value of this expression when evaluated at e^*, so the principal's utility is $U(k^*)$.

Now consider the actual problem, maximizing $EU(\phi,e)$ subject to (5) and (6). Let

$$\phi^*(x,z) = x - k^*. \tag{A10}$$

(Note that ϕ^* depends only on the outcome x.) Then

$$EV(\phi^*,e) = a(e)(\bar{x}(e) - k^*) - v(e). \tag{A11}$$

It is easy to see that $\text{Max}_e\, EV(\phi^*,e) = V^0$ and occurs if $e = e^*$. (For $EV(x - k^*, e^*) = V^0$; and if there were an $\hat{e}$ such that $EV(x - k^*, \hat{e}) > V^0$, then one could find a $k > k^*$ such that $EV(x - k, \hat{e}) = V^0$, contradicting the definition of k^*.) Since e^* and ϕ^* satisfy (5) and (6) and maximize $EU(\phi,e)$ subject only to (5), they certainly maximize $EU(\phi,e)$ subject to (5) and (6). This completes the proof of Proposition 4. Proposition 1 also follows, since a Pareto optimal schedule depending on x and z turns out to depend only on x.

□ **Proof of proposition 5.** Suppose that the fee schedule is a function only of x, say $\phi(x)$, and that it is Pareto optimal. We shall show that this leads to a contradiction by constructing a new fee schedule depending on z as well which makes both the principal and the agent better off. Let $\hat{e}$ be the effort chosen by the agent if the fee schedule is $\phi(x)$. Also, let s be a function of z and x with mean $\bar{s}(x,e)$ given x and e and with $(\partial\bar{s}(x,e)/\partial e)\big|_{e=\hat{e}} > 0$. That is, given any outcome x, s is higher on average if true effort is raised from the level $\hat{e}$. It is clear from our previous remarks that since z conveys information about e, the required s exists. Let $j(x,e)$ be the number of the moment of z, conditional on x and e, which has a nonzero derivative with respect to e. Then s can be taken to be (plus or minus) $z^{j(x,\hat{e})}$. Define the new fee schedule by

$$\omega(x,z) = \phi(x) + \alpha(s - \bar{s}(x,\hat{e})) + \alpha\beta, \tag{A12}$$

where α and β are positive constants to be determined below.

Notice that the new fee schedule ω pays the agent more on average if he continues to take action $\hat{e}$, for under ω he always gets an additional $\alpha\beta$ and the mean of $\alpha(s - \bar{s}(x,\hat{e}))$ is zero. On the other hand, ω gives an additional incentive to the agent to increase e: If e is raised from $\hat{e}$, s will exceed $\bar{s}(x,\hat{e})$ on average, raising the fee on average.

The agent now maximizes over e

$$j(e,\alpha,\beta) = \iint V(\phi(x) + \alpha(s - \bar{s}(x,\hat{e})) + \alpha\beta, e)q(z|x; e)dz\, r(x; e)dx \tag{A13}$$

and the first-order condition

$$j_e(e,\alpha,\beta) = 0 \tag{A14}$$

determines the optimal e as a function of α and β. Let $J(\alpha,\beta) = \max_e j(e,\alpha,\beta)$, so that

$$J_\alpha(\alpha,\beta) = j_e(e,\alpha,\beta)e_\alpha(\alpha,\beta) + j_\alpha(e,\alpha,\beta) = j_\alpha(e,\alpha,\beta). \tag{A15}$$

Since $e(0,\beta) = \hat{e}$,

$$\begin{aligned} J_\alpha(0,\beta) &= j_\alpha(\hat{e},0,\beta) \\ &= \iint (s - \bar{s}(x,\hat{e}) + \beta)V_1(\phi(x),\hat{e})q(z|x;\hat{e})dz\, r(x,\hat{e})dx \\ &= \beta \int V_1(\phi(x),\hat{e})r(x;\hat{e})dx > 0. \end{aligned} \tag{A16}$$

Thus, for any positive β, if α is chosen small enough, the agent will be better off.

To complete the proof, we must show that the principal as well can be made better off. The principal's expected utility as a function of α and β is

$$K(\alpha,\beta) = \iint U(x - \phi(x) - \alpha(s - \bar{s}(x,\hat{e})) - \alpha\beta)q(z|x;e)dz\, r(x;e)dx, \tag{A17}$$

where $e = e(\alpha,\beta)$. We need to show that $K_\alpha(0,\beta) > 0$, if β is appropriately chosen. Now

$$\begin{aligned} K_\alpha(0,\beta) = e_\alpha(0,\beta) \int U(x - \phi(x))r_e(x;\hat{e})dx - \beta \int U'(x - \phi(x))r(x;\hat{e})dx \\ + e_\alpha(0,\beta) \iint U(x - \phi(x))q_e(z|x;\hat{e})dz\, r(x;\hat{e})dx. \end{aligned} \tag{A18}$$

Since $\int q_e(z|x;\hat{e})dz = 0$, the last term is zero. Note that $\int U(x - \phi(x))r_e(x;\hat{e})dx$ is the marginal change under the supposed optimal schedule ϕ in the principal's expected utility if the agent increases e from $\hat{e}$.

Assume first that this term is positive. Given this assumption we need to show that $e_\alpha(0,\beta)$ is greater than some $\delta > 0$, no matter how small β is. Now

$$e_\alpha(0,\beta) = \frac{-j_{e\alpha}(\hat{e},0,\beta)}{j_{ee}(\hat{e},0,\beta)} \tag{A19}$$

and since $j_{ee}(\hat{e},0,\beta) < 0$ (the second-order condition for a regular maximum) and is independent of β, it is enough to show that $j_{e\alpha}(\hat{e},0,\beta)$ is greater than some $\delta > 0$ no matter how small β is. Since

$$\begin{aligned} j_e(e,\alpha,\beta) = \iint V_2(\phi(x) + \alpha(s - \bar{s}(x,e)) + \alpha\beta, e)q(z|x;e)dz\, r(x;e)dx \\ + \iint V(\phi(x) + \ldots, e)q_e(z|x;e)dz\, r(x;e)dx \\ + \iint V(\phi(x) + \ldots, e)q(z|x;e)dz\, r_e(x;e)dx, \end{aligned} \tag{A20}$$

we have

$$\begin{aligned} j_{e\alpha}(e,\alpha,\beta) = \iint (s - \bar{s}(x,e) + \beta)V_{21}(\phi(x) + \ldots, e)q(z|x;e)dz\, r(x;e)dx \\ + \iint (s - \bar{s}(x,e) + \beta)V_1(\phi(x) + \ldots, e)q_e(z|x;e)dz\, r(x;e)dx \\ + \iint (s - \bar{s}(x,e) + \beta)V_1(\phi(x) + \ldots, e)q(z|x;e)dz\, r_e(x;e)dx \end{aligned} \tag{A21}$$

so that

$$j_{e\alpha}(\hat{e},0,\beta) = \beta \int (V_{21}(\phi(x),\hat{e})r(x;\hat{e}) + V_1(\phi(x),\hat{e})r_e(x;\hat{e}))dx$$

$$+ \iint sq_e(z\,|x;\hat{e})dz\; V_1(\phi(x),\hat{e})r(x;\hat{e})dx. \quad \text{(A22)}$$

But since $(\partial\bar{s}(x,e)/\partial e)\big|_{e=\hat{e}} > 0$, we have $\int sq_e(z\,|x;\hat{e})dz > 0$. Therefore $j_{e\alpha}(\hat{e},0,\beta)$ is of the form $K_1 + \beta K_2$ with $K_1 > 0$. Consequently, let $\delta = K_1/2$, so that $j_{e\alpha}(\hat{e},0,\beta) > \delta$ for all small β.

If, on the other hand, $\int U(x - \phi(x))r_e(x;\hat{e})dx$ is negative, the proof just given may be repeated, the only changes being that $\omega(x,z)$ is instead defined as $\phi(x) - \alpha(s - \bar{s}(x,\hat{e})) + \alpha\beta$ and we show that $-e_\alpha(0,\beta)$ is greater than some $\delta > 0$ for all small β.

The case $\int U(x - \phi(x))r_e(x;\hat{e})dx = 0$ is impossible. To demonstrate this we shall show that if equality does hold, a new fee schedule η depending only on x can be constructed so as to make the principal and the agent better off. This will contradict the presumed Pareto optimality of ϕ among fee schedules depending only on x.

Before defining η, we assert that there must exist a perturbation of ϕ which, if effort is held constant (at $\hat{e}$), improves the expected utility of the principal and the agent. Otherwise ϕ would allocate risk in a Pareto optimal way. But we show at the end of the proof that this cannot happen, which justifies our assertion.

Now define a family of new fee schedules by

$$\eta(x,t) = \phi(x) + tp(x), \quad \text{(A23)}$$

where p is a function such that if ϕ is perturbed in the direction p, the principal and the agent are made better off given $\hat{e}$. That is, if

$$EU(t) = \int U(x - \phi(x) - tp(x))r(x;\hat{e})dx \quad \text{(A24)}$$

$$EV(t) = \int V(\phi(x) + tp(x),\, \hat{e})r(x;\hat{e})dx, \quad \text{(A25)}$$

then

$$EU'(0) > 0, \qquad EV'(0) > 0. \quad \text{(A26)}$$

Now let

$$\hat{E}U(t) = \int U(x - \phi(x) - tp(x))r(x;e)dx, \quad \text{(A27)}$$

where e is determined by

$$\max_e \int V(\phi(x) + tp(x),\, e)r(x;e)dx \quad \text{(A28)}$$

and let $\hat{E}V(t)$ be the value of the maximum for (A28). We need to show that $\hat{E}U'(0) > 0$ and $\hat{E}V'(0) > 0$ to establish the contradiction to the optimality of ϕ. But

$$\hat{E}U'(0) = \left(\frac{de}{dt}\right)\left(\int U(x - \phi(x))r_e(x;\hat{e})dx\right) + EU'(0) = EU'(0) > 0, \quad \text{(A29)}$$

since the second factor in parentheses is zero by hypothesis. Similarly, $\hat{E}V'(0) > 0$. Thus, if t is sufficiently small, $\eta(x,t)$ will be Pareto superior to $\phi(x)$.

It remains to prove our claim that ϕ cannot allocate risk in a Pareto optimal way. When the principal is risk neutral, Proposition 2(a) establishes the claim. Thus we may assume that the principal is risk averse. To establish the claim in this case, we will assume that ϕ does allocate the risky x (with density $r(x;\hat{e})$) in a Pareto optimal way, and then show that both the principal and the agent can be made better off.

Under this assumption, note that ϕ must satisfy Borch's equation

$$\gamma V_1(\phi(x),\hat{e}) = U'(x - \phi(x)) \tag{A30}$$

over x for some $\gamma > 0$. Now define a function $h(x)$ by $h(x) = -a < 0$ for $x \leqq x^0$ and $h(x) = b > 0$ for $x > x^0$, where x^0 is such that $0 < \text{Prob}\ (x \leqq x^0) < 1$, given that $e = \hat{e}$. Furthermore, select a and b so that

$$\begin{aligned} 0 &= \int h(x) V_1(\phi(x),\hat{e}) r(x;\hat{e})dx \\ &= -a \int_{x \leqq x^0} V_1(\phi(x),\hat{e}) r(x;\hat{e})dx + b \int_{x > x^0} V_1(\phi(x),\hat{e}) r(x;\hat{e})dx. \end{aligned} \tag{A31}$$

Now consider a new fee schedule $\phi(x) + \alpha h(x) + \alpha\beta$ and, using this schedule, define $j(\alpha,\beta,e)$, $J(\alpha,\beta)$, and $K(\alpha,\beta)$ as before. We have

$$\begin{aligned} J_\alpha(0,\beta) = j_\alpha(0,\beta,\hat{e}) &= \int (\beta + h(x)) V_1(\phi(x),\hat{e}) r(x;\hat{e})dx \\ &= \beta \int V_1(\phi(x),\hat{e}) r(x;\hat{e})dx. \end{aligned} \tag{A32}$$

Thus, for any $\beta > 0$ if α is sufficiently small the agent can be made better off. To complete the proof we must show that for some $\beta > 0$ the principal too can be made better off. Since

$$K(\alpha,\beta) = \int U(x - \phi(x) - \alpha h(x) - \alpha\beta) r(x;e)dx, \tag{A33}$$

where $e = e(\alpha,\beta)$ is chosen by the agent,

$$\begin{aligned} K_\alpha(0,\beta) = &- \int (h(x) + \beta) U'(x - \phi(x)) r(x;\hat{e})dx \\ &+ e_\alpha(0,\beta) \int U(x - \phi(x)) r_e(x;\hat{e})dx. \end{aligned} \tag{A34}$$

But the first term reduces to $-\beta \int U'(x - \phi(x)) r(x;\hat{e})dx$, since by (A30) and (A31), $\int h(x)U'(x - \phi(x))r(x;\hat{e})dx = \gamma \int h(x)V_1(\phi(x),\hat{e})r(x;\hat{e})dx = 0$. Also, by (A30), $x - \phi(x)$ must be strictly increasing in x. Hence $\int U(x - \phi(x))r_e(x;\hat{e})dx > 0$. Moreover, it may be verified from differentiation of the first-order condition $j_e(\alpha,\beta,e) = 0$ that $e_\alpha(0,\beta) > \delta$ for some positive δ no matter how small is β. Thus $K_\alpha(0,\beta) > -\beta \int U'(x - \phi(x))r(x;\hat{e})dx + \delta \int U(x - \phi(x))r_e(x;\hat{e})dx$, which is clearly greater than zero for any positive β sufficiently small. *Q.E.D.*

The proof of Proposition 5 used the assumption that the first-order condition $EV_e(\phi,e) = 0$ determines the agent's choice of e. To see that this assumption is necessary, we shall sketch a situation in which the assumption does not hold and imperfect information about the agent's action is not valuable, although perfect information is.

Suppose that the principal is risk neutral and that the agent is risk averse with $V(w,e) = V(w - e)$, so e is an expenditure. Suppose also that the first-best level of e is $e^* > 0$. Finally, suppose that the optimal ϕ depending only on x is $\phi(x) = k$, a constant. (See footnote 14.) In this case, $e = 0$ and $EV_e(\phi,0) < 0$.

Since $EV_e(\phi,0) < 0$, making the fee schedule depend only slightly on a z that conveys information would not induce the agent to raise e. Only a gross dependence would accomplish this, but it could also result in the imposition of substantial risk. Therefore, an imprecise z might be of no use (even though a perfectly accurate z would be of use, since $e^* > 0$).

References

ARROW, K. "Control in Large Organizations." *Essays in the Theory of Risk Bearing*. Chicago: Markham, 1971a, pp. 223–238.

———. "Exposition of the Theory of Choice under Uncertainty." *Essays in the Theory of Risk Bearing*. Chicago: Markham, 1971b, pp. 44–89.

BORCH, K. "Equilibrium in a Reinsurance Market." *Econometrica*, Vol. 30 (1962), pp. 424–444.

DEGROOT, M. *Optimal Statistical Decisions*. New York: McGraw-Hill, 1970.

HARRIS, M. AND RAVIV, A. "Optimal Incentive Contracts with Imperfect Information." Carnegie Mellon University, Mimeo, 1976.

——— AND ———. "Some Results on Incentive Contracts." *The American Economic Review*, Vol. 68, No. 1 (March 1978a), pp. 20–30.

——— AND ———. "Optimal Incentive Contracts with Imperfect Information." Carnegie Mellon University, mimeo, revised 1978b.

HOLMSTRÖM, B. "Moral Hazard and Observability." *The Bell Journal of Economics*, Vol. 10, No. 1 (Spring 1979), pp. 74–91.

MIRRLEES, J. "Notes on Welfare Economics, Information, and Uncertainty" in M. Balch, D. McFadden, and S. Wu, eds., *Essays on Economic Behavior under Uncertainty*, Amsterdam: North-Holland, 1974, pp. 243–257.

———. "On Moral Hazard and the Theory of Unobservable Behavior." Oxford University, mimeo, 1975.

———. "The Optimal Structure of Incentives and Authority within an Organization." *The Bell Journal of Economics*, Vol. 7, No. 1 (Spring 1976), pp. 105–131.

RAO, C.R. *Linear Statistical Inference and Its Applications*. New York: John Wiley, 1965.

ROSS, S. "The Economic Theory of Agency: The Principal's Problem." *The American Economic Review*, Vol. 63, No. 2 (May 1973), pp. 134–139.

SAVAGE, L. *The Foundations of Statistics*, 2nd rev. ed. New York: Dover, 1972.

SHAVELL, S. "Accidents, Liability, and Insurance." Harvard University, mimeo, 1978.

———. "On Moral Hazard and Insurance." *Quarterly Journal of Economics* (forthcoming).

SPENCE, A.M. AND ZECKHAUSER, R. "Insurance, Information, and Individual Action." *The American Economic Review*, Vol. 61, No. 2 (May 1971), pp. 380–387.

STIGLITZ, J. "Risk Sharing and Incentives in Sharecropping." *Review of Economic Studies*, Vol. 61, No. 2 (April 1974), pp. 219–256.

———. "Incentives, Risk and Information: Notes Toward a Theory of Hierarchy." *The Bell Journal of Economics*, Vol. 6, No. 2 (Autumn 1975), pp. 552–579.