Econometrica, Vol. 51, No. 1 (January, 1983)

AN ANALYSIS OF THE PRINCIPAL-AGENT PROBLEM

BY SANFORD J. GROSSMAN AND OLIVER D. HART[1]

Most analyses of the principal-agent problem assume that the principal chooses an incentive scheme to maximize expected utility subject to the agent's utility being at a stationary point. An important paper of Mirrlees has shown that this approach is generally invalid. We present an alternative procedure. If the agent's preferences over income lotteries are independent of action, we show that the optimal way of implementing an action by the agent can be found by solving a convex programming problem. We use this to characterize the optimal incentive scheme and to analyze the determinants of the seriousness of an incentive problem.

1. INTRODUCTION

IT HAS BEEN RECOGNIZED for some time that, in the presence of moral hazard, market allocations under uncertainty will not be unconstrained Pareto optimal (see Arrow [**1**], Pauly [**13**]). It is only relatively recently, however, that economists have begun to undertake a systematic analysis of the properties of the second-best allocations which will arise under these conditions. Much of this analysis has been concerned with what has become known as the principal-agent problem. Consider two individuals who operate in an uncertain environment and for whom risk sharing is desirable. Suppose that one of the individuals (known as the agent) is to take an action which the other individual (known as the principal) cannot observe. Assume that this action affects the total amount of consumption or money which is available to be divided between the two individuals. In general, the action which is optimal for the agent will depend on the extent of risk sharing between the principal and the agent. The question is: What is the optimal degree of risk sharing, given this dependence?

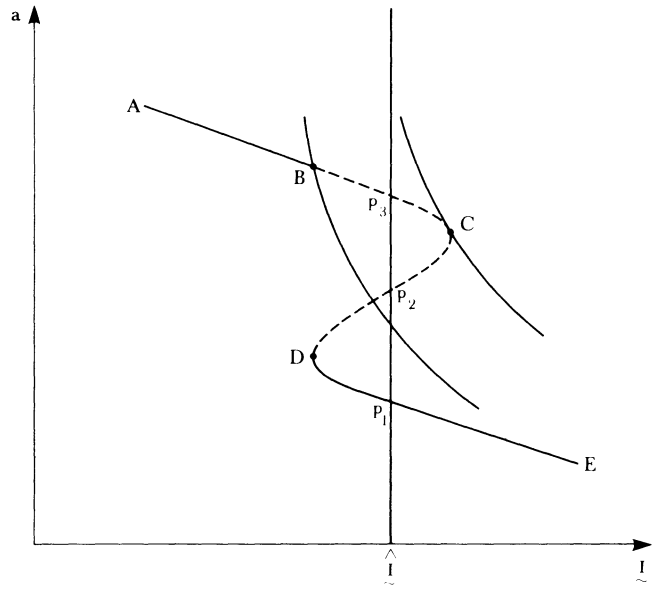
Particular applications of the principal-agent problem have been made to the case of an insurer who cannot observe the level of care taken by the person being insured; to the case of a landlord who cannot observe the input decision of a tenant farmer (sharecropping); and to the case of an owner of a firm who cannot observe the effort level of a manager or worker.[2]

Although considerable progress has been made in the recent literature towards understanding and solving the principal-agent problem (see, in particular, Harris and Raviv [**6**], Holmstrom [**7**], Mirrlees [**10**, **11**, **12**], Shavell [**19**, **20**], as well as the other references in footnote 2), the mathematical approach which has been adopted in most of this literature is unsatisfactory. The procedure usually followed is to suppose that the principal chooses the risk-sharing contract, or incentive scheme, to maximize his expected utility subject to the constraints that

[1]Support from the U.K. Social Science Research Council and NSF Grant No. SOC70-13429 is gratefully acknowledged. We would like to thank Bengt Holmstrom, Mark Machina, Andreu Mas-Colell, and Jim Mirrlees for helpful comments.

[2]These and other applications are discussed in a number of recent papers. See, for example, Harris and Raviv [**6**], Holmstrom [**7**], Mirrlees [**10**, **11**, **12**], Radner [**15**], Ross [**17**], Rubinstein and Yaari [**18**], Shavell [**19**, **20**], Spence and Zeckhauser [**21**], Stiglitz [**22**], and Zeckhauser [**24**].

G. Dionne et al. (eds.), *Foundations of Insurance Economics*



For a given $\underset{\sim}{I}$ the agent strictly prefers lower actions

FIGURE 1.

(a) the agent's expected utility is no lower than some pre-specified level; (b) the agent's utility is at a stationary point, i.e., the agent satisfies his first-order conditions with respect to the choice of action. That is, the agent's second-order conditions (and the condition that the agent should be at a global rather than a local maximum) are ignored. Mirrlees [**10**], however, in an important paper, has shown that this procedure is generally invalid unless, at the optimum, the solution to the agent's maximum problem is unique. In the absence of uniqueness (and it is difficult to guarantee uniqueness in advance), the first-order conditions derived by the above procedure are not even necessary conditions for the optimality of the risk-sharing contract.[3]

[3]The reason for this can be seen quite easily in Figure 1 (we are grateful to Andreu Mas-Colell for suggesting the use of this figure). On the horizontal axis, $\boldsymbol{I}$ represents the agent's incentive scheme and on the vertical axis a represents the agent's action. The curve $ABCDE$ is the locus of pairs of actions and incentive schemes which satisfy the agent's first order conditions, i.e., given $\boldsymbol{I}$ the agent's utility is at a stationary point. Of these points, only those lying on the segments AB and DE represent global maxima for the agent, e.g. given the incentive scheme $\hat{\boldsymbol{I}}$ the agent's optimal action is at p_1, not at p_2 or p_3. Indifference curves—in terms of a and $\boldsymbol{I}$—are drawn for the principal (C is on a higher curve than B). The true feasible set for the principal are the segments AB and DE and the optimal outcome for the principal is therefore B. However, B does not satisfy the first order conditions of the problem: maximize the principal's utility subject to $(a, \boldsymbol{I})$ lying on $ABCDE$, i.e., subject to $(a, \boldsymbol{I})$ satisfying the agent's first order conditions (the solution to this problem is at C). In other words, B does not satisfy the necessary conditions for optimality of the problem which has been studied in much of the literature. Note finally that perturbing Figure 1 slightly does not alter this conclusion.

The purpose of this paper is to develop a method for analyzing the principal-agent problem which avoids the difficulties of the "first-order condition" approach.[4] Our approach is to break the principal's problem up into a computation of the costs and benefits of the different actions taken by the agent. For each action, we consider the incentive scheme which minimizes the (expected) cost of getting the agent to choose that action. We show that, under the assumption that the agent's preferences over income lotteries are independent of the action he takes, this cost minimization problem is a fairly straightforward convex programming problem. An analysis of these convex problems as the agent's action varies yields a number of results about the form of the optimal incentive scheme. We will also be able to analyze what factors determine how serious a particular incentive problem is; i.e., how great the loss is to the principal from having to operate in a second-best situation where the agent's action cannot be observed relative to a first-best situation where it can be observed.

The assumption that the agent's preferences over income lotteries are independent of action is a strong one. Yet it seems a natural starting point for an analysis of the principal-agent problem. Special cases of this assumption occur when the agent's utility function is additively or multiplicatively separable in action and reward. One or other of these cases is typically assumed in most of the literature. In Section 6 we discuss briefly the prospects for the non-independence case.

In addition to providing greater rigor, the costs versus benefits approach also provides a clear separation of the two distinct roles the agent's output plays in the principal-agent problem. On the one hand, the agent's output contributes positively to the principal's consumption, so the principal desires a high output. On the other hand, the agent's output is a signal to the principal about the agent's level of effort. This informational role may be in conflict with the consumption role. For example, there may be a moderate output level which is achieved when the agent takes low effort levels and never occurs at other effort levels. If the agent is penalized whenever this moderate output occurs, then he is discouraged from taking these low effort actions. However, there may be lower output levels which have some chance of occurring regardless of the agent's action. To encourage the agent to take high effort levels, it is then optimal to pay the agent more in low output states than in moderate output states, even though the principal prefers moderate output levels to low output levels.

The dual role of output makes it difficult to obtain conditions which ensure even elementary properties of the incentive scheme, such as monotonicity. In Section 3, sufficient conditions for monotonicity are given. It is also shown in this section that a monotone likelihood ratio condition, which the "first-order condition" approach suggests is a guarantee of monotonicity, must be strengthened once we take into account the possibility that the agent's action is not unique at the optimal incentive scheme.

The paper is organized as follows. In Section 2, we show how the principal's optimization problem can be decomposed into a costs versus benefits problem.

[4] Mirrlees [12] has identified a class of cases where the "first-order condition" approach *is* valid. We will consider this class in Section 3.

In Section 3, we use our approach to analyze the monotonicity and progressivity of the optimal incentive scheme. In Section 4, we give a simple algorithm for computing an optimal incentive scheme when there are only two outcomes associated with the agent's actions. In Section 5, we analyze the effects of risk aversion and information quality on the incentive problem. Finally, in Section 6 we consider some extensions of the analysis.

2. STATEMENT OF THE PROBLEM

The application of the principal-agent problem that we will consider is to the case of the owner of a firm who delegates the running of the firm to a manager. The owner is the principal and the manager the agent. The owner is assumed not to be able to monitor the manager's actions. The owner does, however, observe the outcome of these actions, which we will take to be the firm's profit. It is assumed that the firm's profit depends on the manager's actions, but also on other factors which are outside the manager's control—we model these as a random component. Thus, in particular, if the firm does well, it will not generally be clear to the owner whether this is because the manager has worked well or whether it is because he has been lucky.[5]

We will simplify matters by assuming that there are only finitely many possible gross profit levels for the firm, denoted $q_1, \ldots, q_n$, where $q_1 < q_2 < \cdots < q_n$. We will assume that the principal is interested only in the firm's net profit, i.e. gross profit minus the payment to the manager. We will also assume that the principal is risk neutral—our methods of analysis can, however, be applied to the case where the principal is risk averse (see Remark 3 and Section 6).

Let A be the set of actions available to the manager. We will assume that A is a non-empty, compact subset of a finite dimensional Euclidean space. Let $S = \{x \in R^n \mid x \geq 0, \sum_{i=1}^n x_i = 1\}$. We assume that there is a continuous function $\pi : A \to S$, where $\pi(a) = (\pi_1(a), \ldots, \pi_n(a))$ gives the probabilities of the n outcomes $q_1, \ldots, q_n$ if action a is selected. It is assumed that, when the agent chooses $a \in A$, he knows the probability function π but not the outcome which will result from his action. We assume that the agent has a von Neumann–Morgenstern utility function $U(a, I)$ which depends both on his action a and his remuneration I from the principal. We include a as an argument in order to capture the idea that the agent dislikes working hard, taking care, etc.

The crucial assumption that we will make about the form of $U(a, I)$ is:

ASSUMPTION A1: $U(a, I)$ can be written as $G(a) + K(a)V(I)$, where (1) V is a real-valued, continuous, strictly increasing, concave function defined on some open interval $\mathcal{I} = (\underline{I}, \infty)$ of the real line; (2) $\lim_{I \to \underline{I}} V(I) = -\infty$; (3) G, K are

[5] The assumption that the principal cannot monitor the agent's actions at all may in some cases be rather extreme. For a discussion of the implications of the existence of imperfect monitoring opportunities, see Harris and Raviv [6], Holmstrom [7] and Shavell [19, 20]. See also Remark 4 in Section 2.

real-valued, continuous functions defined on A and K is strictly positive; (4) for all $a_1, a_2 \in A$ and $I, \hat{I} \in \mathcal{I}$, $G(a_1) + K(a_1)V(I) \geq G(a_2) + K(a_2)V(I) \Rightarrow G(a_1) + K(a_1)V(\hat{I}) \geq G(a_2) + K(a_2)V(\hat{I})$.

In the above, we allow for the case $\underline{I} = -\infty$.

The main part of Assumption A1 has a simple ordinal interpretation. Assumption A1 implies that the agent's preferences over income lotteries are independent of his action (Assumption A1(1) tells us also that these preferences exhibit risk aversion). The converse can also be shown to be true: if the agent's preferences over income lotteries are independent of a, then U can be written as $G(a) + K(a)V(I)$ for some functions G, K, V (for a proof, see Keeney [**8**]). Note that Assumption A1 does not imply that the agent's preferences for *action* lotteries are independent of income. We will insist, however, that the agent's ranking over *perfectly certain* actions is independent of income—this is condition (4) of Assumption A1.

Note that if $K(a)$ is not constant then (2) and (4) imply that $V(I)$ must be bounded from above. Further if it is also the case that $G(a) \equiv 0$, then $V(I)$ must be non-positive everywhere.

Two special cases of Assumption A1 occur when $K(a) =$ constant, i.e. U is additively separable in a and I, and when $G(a) = 0$, i.e. U is multiplicatively separable in a and I. In these cases the agent's preferences over action lotteries *are* independent of income, as well as preferences over income lotteries being independent of action.[6]

An interesting special case of multiplicative separability is when $V(I) = -e^{-kI}$, $K(a) = e^{ka}$ and A is a subset of the real line. Then $U(a, I) = -e^{-k(I-a)}$; i.e., effort appears just as negative income.

In the "first-best" situation where the principal can observe a, it is optimal for him to pay the agent according to the action he chooses. Let $\overline{U}$ be the agent's reservation price, i.e. the expected level of utility he can achieve by working elsewhere, and let $\mathcal{U} = V(\mathcal{I}) = \{v \mid v = V(I)$ for some $I \in \mathcal{I}\}$. We make the following assumption.

ASSUMPTION A2: $[\overline{U} - G(a)]/K(a) \in \mathcal{U}$ for all $a \in A$.

DEFINITION: Let $C_{FB} : A \to R$ be defined by $C_{FB}(a) = h([\overline{U} - G(a)]/K(a))$, where $h \equiv V^{-1}$.

Here C_{FB} stands for first-best cost. $C_{FB}(a)$ is simply the agent's reservation price for picking action a. To get the agent to pick $a \in A$ in the first-best

[6]The converse is also true: if preferences over action lotteries are independent of income as well as preferences over income lotteries being independent of action, then U is additively or multiplicatively separable (see Keeney [**8**] or Pollak [**14**]).

situation, the principal will offer him the following contract: I will pay you $C_{FB}(a)$ if you choose a and $\tilde{I}$ otherwise, where $\tilde{I}$ is very close to $\underline{I}$.

DEFINITION: Let $B: A \to R$ be defined by $B(a) = \sum_{i=1}^{n} \pi_i(a) q_i$. $B(a)$ is the expected benefit to the principal from getting the agent to pick a.

DEFINITION: A first-best optimal action is one which maximizes $B(a) - C_{FB}(a)$ on A.

The function C_{FB} induces a complete ordering on $A: a \gtrsim a'$ if and only if $C_{FB}(a) \geq C_{FB}(a')$. For obvious reasons we will refer to actions with higher $C_{FB}(a)$'s as costlier actions. It is easy to show, in view of Assumption A1(4), that $C_{FB}(a) \geq C_{FB}(a') \Leftrightarrow G(a) + K(a)v \leq G(a') + K(a')v$ for all $v \in \mathfrak{V} \Leftrightarrow G(a) + K(a)v \leq G(a') + K(a')v$ for some $v \in \mathfrak{V}$. This in turn implies that the ordering $\gtrsim$ is independent of $\bar{U}$. In the second-best situation where a is not observed by the principal, it is not possible to make the agent's remuneration depend on a. Instead, the principal will pay the agent according to the *outcome* of his action, i.e. according to the firm's profit. An incentive scheme is therefore an n-dimensional vector $\boldsymbol{I} = (I_1, I_2, \ldots, I_n) \in \mathcal{I}^n$, where I_i is the agent's remuneration in the event that the firm's profit is q_i. Given the incentive scheme $\boldsymbol{I}$, the agent will choose $a \in A$ to maximize $\sum_{i=1}^{n} \pi_i(a) U(a, I_i)$.

We will assume that the principal knows the agent's utility function $U(a, I)$, the set A and the function $\pi: A \to S$. In other words, the principal is fully informed about the agent and about the firm's production possibilities. The incentive problem which we will study therefore arises entirely because the principal cannot monitor the agent's actions.[7]

The principal's problem can be described as follows. Let F be the set of pairs of incentive schemes $\boldsymbol{I}^*$ and actions a^* such that, under $\boldsymbol{I}^*$, the agent will be willing to work for the principal and will find it optimal to choose a^*, i.e. $\max_{a \in A} \sum_{i=1}^{n} \pi_i(a) U(a, I_i^*) = \sum_{i=1}^{n} \pi_i(a^*) U(a^*, I_i^*) \geq \bar{U}$. Then the principal chooses $(\boldsymbol{I}, a) \in F$ to maximize $\sum_{i=1}^{n} \pi_i(a)(q_i - I_i)$. It simplifies matter considerably if we break this problem up into two parts. We consider first, given that the principal wishes to implement a^*, the least cost way of achieving this. We then consider which a^* should be implemented. Thus, to begin, suppose that the principal wishes the agent to pick a particular action $a^* \in A$. To find the least (expected) cost way of achieving this, the principal must solve the following

[7] This distinguishes our study from the literature on incentive compatibility; see, e.g., the recent *Review of Economic Studies* symposium [**16**]. The incentive compatibility literature has been concerned with incentive problems arising from differences in information between individuals rather than with those arising from monitoring problems. In cases of differential information, there is a role for an exchange of information through messages, whereas in the model we study messages would serve no purpose.

problem:

$$\text{(2.1)} \qquad \text{Choose } I_1, \ldots, I_n \text{ to minimize } \sum_{i=1}^{n} \pi_i(a^*) I_i$$

$$\text{subject to} \quad \sum_{i=1}^{n} \pi_i(a^*) U(a^*, I_i) \geq \bar{U},$$

$$\sum_{i=1}^{n} \pi_i(a^*) U(a^*, I_i) \geq \sum_{i=1}^{n} \pi_i(a) U(a, I_i) \qquad \text{for all} \quad a \in A,$$

$$I_i \in \mathcal{I} \qquad \text{for all } i.$$

This problem can be simplified considerably in view of Assumption A1. It will be convenient to regard $v_1 = V(I_1), \ldots, v_n = V(I_n)$ as the principal's control variables. Recall that $\mathcal{U} = V(\mathcal{I}) = \{v \mid v = V(I) \text{ for some } I \in \mathcal{I}\}$. By Assumption A1, $\mathcal{U}$ is an interval of the real line $(-\infty, \bar{v})$. Thus we may rewrite (2.1) as follows:

$$\text{(2.2)} \qquad \text{Choose } v_1, \ldots, v_n \text{ to minimize } \sum_{i=1}^{n} \pi_i(a^*) h(v_i)$$

$$\text{subject to} \quad G(a^*) + K(a^*)\left(\sum_{i=1}^{n} \pi_i(a^*) v_i\right) \geq G(a) + K(a)\left(\sum_{i=1}^{n} \pi_i(a) v_i\right)$$

$$\text{for all} \quad a \in A,$$

$$G(a^*) + K(a^*)\left(\sum_{k=1}^{n} \pi_i(a^*) v_i\right) \geq \bar{U},$$

$$v_i \in \mathcal{U} \qquad \text{for all } i,$$

where $h \equiv V^{-1}$.

The important point to realize is that the constraints in (2.2) are linear in the v_i's. Furthermore, V concave implies h convex, and so the objective function is convex in the v_i's. Thus (2.2) is a rather simple optimization problem: minimize a convex function subject to (a possibly infinite number of) linear constraints. In particular, when A is a finite set, the Kuhn–Tucker theorem yields necessary and sufficient conditions for optimality. These will be analyzed later.

It is important to realize that, in the absence of Assumption A1, it is not generally possible to convert (2.1) into a convex problem in this way.

DEFINITION: If $\boldsymbol{I} = (I_1, \ldots, I_n)$ satisfies the constraints in (2.1) or $\boldsymbol{v} = (v_1, \ldots, v_n)$ satisfies the constraints in (2.2), we will say that $\boldsymbol{I}$ or $\boldsymbol{v}$ *implements* action a^*. (We are assuming here that if the agent is indifferent between two actions, he will choose the one preferred by the principal.)

Consider the set of $\mathbf{v}$'s which implement a^*. For some a^*, this set may be empty, in which case action a^* cannot be implemented by the principal at any cost. If the set is non-empty, then, since h is convex,

$$\sum_{i=1}^{n} \pi_i(a^*)h(v_i) \geq h\left(\sum_{i=1}^{n} \pi_i(a^*)v_i\right) \geq h\left(\frac{\bar{U} - G(a^*)}{K(a^*)}\right)$$

by (2.2), and so the principal's objective function is bounded below on this set. Let $C(a^*)$ be the greatest lower bound of $\sum_{i=1}^{n} \pi_i(a^*)h(v_i)$ on this set.

DEFINITION: Let $C(a^*) = \inf\{\sum_{i=1}^{n} \pi_i(a^*)h(v_i) \mid \mathbf{v} = (v_1, \ldots, v_n)$ implements $a^*\}$ if the constraint set in (2.2) is non-empty. In the case where the constraint set of (2.2) is empty, write $C(a^*) = \infty$. This defines the second-best cost function $C: A \to Ru\{\infty\}$.

The above constitutes the first step(s) of the principal's optimization problem: for each $a \in A$, compute $C(a)$. The second step is to choose which action to implement, i.e. to choose $a \in A$ to maximize $B(a) - C(a)$. This second problem will not generally be a convex problem. This is because even if $B(a)$ is concave in a, $C(a)$ will not generally be convex. Fortunately, a significant amount of information about the form of the optimal incentive scheme can be obtained by studying the first step alone.

DEFINITION: A second-best optimal action $\hat{a}$ is one which maximizes $B(a) - C(a)$ on A. A second-best optimal incentive scheme $\hat{\mathbf{I}}$ is one that implements a second-best optimal action $\hat{a}$ at least expected cost, i.e. $\sum_{i=1}^{n} \pi_i(\hat{a})\hat{I}_i = C(\hat{a})$.

Note that for a second-best optimal incentive scheme to exist, the greatest lower bound in the definition of $C(a)$ must actually be achieved. In order to establish the existence of a second-best optimal action and a second-best optimal incentive scheme, we need a further assumption.

ASSUMPTION A3: For all $a \in A$ and $i = 1, \ldots, n$, $\pi_i(a) > 0$.

Since there are only finitely many possible profit levels, Assumption A3 implies that $\pi_i(a)$ is bounded away from zero. Hence Assumption A3 rules out cases studied by Mirrlees [12] in which an optimum can be approached but not achieved by imposing higher and higher penalties on the agent which occur with smaller and smaller probability if the agent chooses the right action.

PROPOSITION 1: *Assume* A1–A3. *Then there exists a second-best optimal action* $\hat{a}$ *and a second-best optimal incentive scheme* $\hat{\mathbf{I}}$.

PROOF: It is helpful to split the proof up into two parts. Consider first the case where V is linear. Then it is easy to see that the principal can do as well in the second-best as in the first-best where the agent can be monitored. For let a^* maximize $B(a) - C_{FB}(a)$ on A. Let the principal offer the agent the incentive scheme $I_i = q_i - t$, where $t = B(a^*) - C_{FB}(a^*)$. Then the principal's profit will be $B(a^*) - C_{FB}(a^*)$ whatever the agent does. On the other hand, by picking $a = a^*$, the agent can obtain expected utility $\bar{U}$. Hence Proposition 1 certainly holds when V is linear.

On the other hand, suppose V is not linear. We show first, that, if the constraint set is nonempty for an action $a^* \in A$, then problem (2.2) has a solution, i.e. $\sum_{i=1}^n \pi_i(a^*)h(v_i)$ achieves its greatest lower bound $C(a^*)$. Note that $\sum_{i=1}^n \pi_i(a^*)v_i$ is bounded below on the constraint set of (2.2). It therefore follows from a result of Bertsekas [2] that unbounded sequences in the constraint set make $\sum_{i=1}^n \pi_i(a^*)h(v_i)$ tend to infinity (roughly because the variance of the $v_i \to \infty$ while their mean is bounded below, and h is convex and nonlinear—Assumption A3 is important here). Hence, we can artificially bound the constraint set. Since the constraint set is closed, the existence of a minimum therefore follows from Weierstrass' theorem.

We show next that $C(a)$ is a lower semicontinuous function of a. If A is finite, then any function defined on A is continuous and hence lower semicontinuous. Assume therefore that A is not finite. Let (a_r) be a sequence of points in A converging to a. Assume without loss of generality (w.l.o.g.) that $C(a_r) \to k$. Then, if $k = \infty$, we certainly have $C(a) \leq \lim_{r\to\infty} C(a_r)$. Suppose therefore that $k < \infty$. Let $(I_1^r, \ldots, I_n^r)$ be the solution to (2.1) when $a^* = a_r$. Then Bertsekas' result together with Assumption A3 shows that the sequence $((I_1^r, \ldots, I_n^r))$ is bounded (otherwise $C(a_r) \to \infty$). Let $(I_1, \ldots, I_n)$ be a limit point. Then clearly $(I_1, \ldots, I_n)$ implements a and so $C(a) \leq \sum_{i=1}^n \pi_i(a) I_i = \lim_{r\to\infty} C(a_r)$. This proves lower semicontinuity.

Given that $C(a)$ is lower semicontinuous and A is compact, it follows from Weierstrass' theorem that $\max_{a\in A}[B(a) - C(a)]$ has a solution, as long as $C(a)$ is finite for some $a \in A$. To prove this last part, we show that $C(a^*) = C_{FB}(a^*)$ if a^* minimizes $C_{FB}(a)$ on A. To see this, note that the a^* which minimizes $C_{FB}(a)$ can be implemented by setting $I_i = C_{FB}(a^*)$ for all i.

We have thus established the existence of a second-best optimal action, $\hat{a}$, when V is nonlinear. Since we have also shown that (2.2) has a solution as long as the constraint set is non-empty and V is nonlinear, this establishes the existence of a second-best optimal incentive scheme. *Q.E.D.*

It is interesting to ask whether the constraint that the agent's expected utility be greater than or equal to $\bar{U}$ is binding at a second-best optimum. The answer is no in general, i.e. for incentive reasons it may pay the principal to choose an incentive scheme which gives the agent an expected utility in excess of $\bar{U}$. One case where this will not happen is when the agent's utility function is additively or multiplicatively separable in action and reward:

PROPOSITION 2: *Assume* A1, A2, *and either* $K(a)$ *is a constant function on* A *or* $G(a) = 0$ *for all* $a \in A$. *Let* $\hat{a}$ *be a second-best optimal action and* $\hat{\mathbf{I}}$ *a second-best optimal incentive scheme which implements* $\hat{a}$. *Then* $\sum_{i=1}^{n} \pi_i(\hat{a}) U(\hat{a}, \hat{I}_i) = \overline{U}$.

PROOF: Suppose not. Write $\hat{v}_i = V(\hat{I}_i)$. Then $G(\hat{a}) + K(\hat{a})(\sum_{i=1} \pi_i(\hat{a})\hat{v}_i) > \overline{U}$ in (2.2). But it is clear that the principal's costs can be reduced and all the constraints of (2.2) will still be satisfied if we replace $\hat{v}_i$ by $(\hat{v}_i - \epsilon)$ for all i in the additively separable case and by $v_i(1 + \epsilon)$ for all i in the multiplicatively separable case where $\epsilon > 0$ is small (note that in the multiplicatively separable case, it follows from (2)–(4) of Assumption A1 that $V(I) < 0$ for all $I \in \mathcal{I}$, and so $\hat{v}_i < 0$). In other words, $\hat{a}$ can be implemented at lower expected cost, which contradicts the fact that we are at a second-best optimum. *Q.E.D.*

REMARK 1: The proof of Proposition 1 establishes that $C(a^*) = C_{FB}(a^*)$ if a^* minimizes $C_{FB}(a)$ on A. This is a reflection of the fact that there is no trade-off between risk sharing and incentives when the action to be implemented is a cost-minimizing one (i.e. involves the agent in minimum "effort").

REMARK 2: In general, there may be more than one second-best optimal action and more than one second-best optimal incentive scheme. It is clear from (2.2), however, that, if V is strictly concave, there is a unique second-best optimal incentive scheme which implements any particular second-best optimal action.

DEFINITION: Let $L = \max_{a \in A}(B(a) - C_{FB}(a)) - \sup_{a \in A}(B(a) - C(a))$ be the difference between the principal's expected profit in the first-best and second-best situations.

L represents the loss which the principal incurs as a result of being unable to observe the agent's action (we write $\sup(B(a) - C(a))$ rather than $\max(B(a) - C(a))$ to cover cases where the assumptions of Proposition 1 do not hold). Proposition 3 shows that, while there are some special cases in which $L = 0$, in general $L > 0$.

PROPOSITION 3: *Assume* A1 *and* A2. *Then*: (1) $C(a) \geq C_{FB}(a)$ *for all* $a \in A$, *which implies that* $L \geq 0$. (2) *If* V *is linear*, $L = 0$. (3) *If there exists a first-best optimal action* $a^* \in A$ *satisfying*: *for each* i, $\pi_i(a^*) > 0 \Rightarrow \pi_i(a) = 0$ *for all* $a \in A$, $a \neq a^*$, *then* $L = 0$. (4) *If* A *is a finite set and there is a first-best optimal action* a^* *which satisfies*: *for some* i, $\pi_i(a^*) = 0$ *and* $\pi_i(a) > 0$ *for all* $a \in A$, $a \neq a^*$, *then* $L = 0$. (5) *If there is a first-best optimal action* $a^* \in A$ *which minimizes* $C_{FB}(a)$ *on* A, $L = 0$. (6) *If Assumption A3 holds, every maximizer* $\tilde{a}$ *of* $B(a) - C_{FB}(a)$ *on* A *satisfies* $C_{FB}(\tilde{a}) > \min_{a \in A} C_{FB}(a)$, *and* V *is strictly concave, then* $L > 0$.

PROOF: (1) is obvious since anything which is second-best feasible is also first-best feasible. (2) follows from the first part of the proof of Proposition 1. (5) follows from the proof of Proposition 1 (see also Remark 1). (3) and (4) follow from the fact that a^* can be implemented by offering the agent $I_i = C_{FB}(a^*)$ for those i such that $\pi_i(a^*) > 0$ and I close to $\underline{I}$ otherwise.

To prove (6), note that, if V is strictly concave,

$$G(a^*) + K(a^*) \sum_{i=1}^{n} \pi_i(a^*) V(I_i) \geq \bar{U}$$

implies

$$\begin{aligned} C(a^*) &= \sum_{i=1}^{n} \pi_i(a^*) h(V(I_i)) \\ &> h\left(\sum_{i=1}^{n} \pi_i(a^*) V(I_i) \right) \geq h\left(\left(\bar{U} - G(a^*) \right) / K(a^*) \right) \\ &= C_{FB}(a^*) \end{aligned}$$

unless $I_i =$ constant with probability 1. But, since $\pi_i(a^*) > 0$ for all i, $I_i =$ constant with probability $1 \Rightarrow I_i$ is independent of i. However, in this case, the constraints of problem (2.2) imply that $C_{FB}(a)$ is minimized at a^*. *Q.E.D.*

Most of Proposition 3 is well known. Proposition 3(2) and (6) can be understood as follows. In the first-best situation, if the agent is strictly risk averse, the principal bears all the risk and the agent bears none. In the second best situation, this is generally undesirable. For if the agent is completely protected from risk, then he has no incentive to work hard; i.e., he will choose $a \in A$ to minimize $C_{FB}(a)$. Hence the second-best situation is strictly worse from a welfare point of view than the first-best situation. The exception is when the agent is risk neutral, in which case it is optimal both from a risk sharing and an incentive point of view for him to bear all the risk, or when the first-best optimal action is cost minimizing.

In the case of Proposition 3(3) and 3(4), a scheme in which the agent is penalized very heavily if certain outcomes occur can be used to achieve the first best. This relates to results obtained in Mirrlees [**12**].

REMARK 3: We have assumed that the principal is risk neutral. Our analysis generalizes to the case where the principal is risk averse, however. In this case, instead of choosing v to minimize $\sum \pi_i(a^*) h(v_i)$ in problem (2.2), we choose v to maximize $\sum \pi_i(a^*) U_p(q_i - h(v_i))$, where U_p is the principal's utility function. Note that (2.2) is still a convex problem. Although we can no longer analyze costs and benefits separately, we can, for each $a^* \in A$, define a net benefit function $\max_v \sum \pi_i(a^*) U_p(q_i - h(v_i))$. An optimal action for the principal is now one that maximizes net benefits. See also Section 6 on this.

REMARK 4: We have taken the outcomes observed by the principal to be profit levels. Our analysis generalizes, however, to the case where the outcomes are more complicated objects, such as vectors of profits, sales, etc., or to the case where profits are not observed at all but something else is (see, e.g., Mirrlees [**11**]). The important point to realize is that profit does not appear in the cost

minimization problem (2.1) or (2.2). Thus, if the principal observes the realizations of a signal $\tilde{\theta}$, then I_i refers to the payment to the agent when $\tilde{\theta} = \theta_i$. Let $\hat{C}(a, \tilde{\theta})$ be the cost of implementing a when the information structure is $\tilde{\theta}$ (e.g. if $\tilde{\theta}$ reveals a exactly, then $\hat{C}(a, \tilde{\theta}) = C_{FB}(a)$). Note that if the distribution of output is generated by a production function $f(a, \tilde{w})$, such that the marginal distribution of $\tilde{w}$ is independent of the information structure, then $B(a) = Ef(a, \tilde{w}) = E[E[f(a, \tilde{w}) | \theta]]$ is independent of the information structure, given a. It follows that the effect of changes in the information structure is summarized by the way that $C(a, \tilde{\theta})$ changes when the information structure changes. As will be seen in Section 5, this is quite easy to analyze.

3. SOME CHARACTERISTICS OF OPTIMAL INCENTIVE SCHEMES

It is of interest to know whether the optimal incentive scheme is monotone increasing (i.e., whether the agent is paid more when a higher output is observed) and whether the scheme is progressive (i.e., whether the marginal benefit to the agent of increased output is decreasing in output). These questions are quite difficult to answer because of the informational role of output. As we noted in the introduction, the agent may be given a low income at intermediate levels of output in order to discourage particular effort levels. Nevertheless, some general results about the shape of optimal schemes can be established. We begin with the following lemma.

LEMMA 1: *Assume* A1–A3. *Let* $(I_i)_{i=1}^n, (I_i')_{i=1}^n$ *be incentive schemes which cause* a *and* a' *to be optimal choices for the agent, respectively, and minimize the respective costs (i.e.* (2.1) *or* (2.2) *is solved). Let* $v_i = V(I_i)$ *and* $v_i' = V(I_i')$. *Then, if* $G(a) + K(a)(\sum_{i=1}^n \pi_i(a) v_i) = G(a') + K(a')(\sum_{i=1} \pi_i(a') v_i')$, *i.e. the agent's expected utility is the same under both schemes, we must have*

$$\sum_i \left[\pi_i(a') - \pi_i(a)\right](v_i' - v_i) \geq 0. \tag{3.1}$$

PROOF: From (2.2) and the assumption that the agent's expected utility is the same, we have

$$G(a') + K(a')\left(\sum_{i=1}^n \pi_i(a') v_i\right) \leq G(a) + K(a)\left(\sum_{i=1}^n \pi_i(a) v_i\right)$$

$$= G(a') + K(a')\left(\sum_{i=1}^n \pi_i(a') v_i'\right),$$

$$G(a) + K(a)\left(\sum_{i=1}^n \pi_i(a) v_i'\right) \leq G(a') + K(a')\left(\sum_{i=1}^n \pi_i(a') v_i'\right)$$

$$= G(a) + K(a)\left(\sum_{i=1}^n \pi_i(a) v_i\right).$$

It follows from the first of these that $\sum_{i=1}^{n}\pi_i(a')(v_i' - v_i) \geq 0$ and from the second that $\sum_{i=1}^{n}\pi_i(a)(v_i - v_i') \geq 0$ (since $K(a) > 0$ by Assumption A1(3)). Adding yields (3.1). *Q.E.D.*

We now use Lemma 1 to show that an optimal incentive scheme will have the property that the principal's and agent's returns are positive related over some range of output levels; i.e., it is not optimal to have, for all output levels q_i, $q_j : I_i > I_j \Rightarrow q_i - I_i < q_j - I_j$. The proof proceeds by showing that, if the principal's and agent's payments are negatively related, then a twist in the incentive schedule which raises the agent's payment in high return states for the principal and lowers it in low return states for the principal can make the principal better off. The reason is that such a twist will be good for incentives since it gets the agent to put more probability weight on states yielding the principal a high return, and it is also good for risk-sharing since it raises the agent's return in low return states for the agent and lowers the agent's return in high return states for the agent. Since the incentive and risk-sharing effects reinforce each other, the principal is made better off.

In order to bring about both the incentive and risk-sharing effects, the twist in the incentive scheme must be chosen carefully. It is for this reason that the proof of the next proposition may seem rather complicated at first sight.

PROPOSITION 4: *Assume* A1–A3 *and V strictly concave. Let* $(I_1, \ldots, I_n)$ *be a second-best optimal incentive scheme. Then the following cannot be true:* $I_i > I_j \Rightarrow q_i - I_i \leq q_j - I_j$ *for all* $1 \leq i, j \leq n$ *and for some* i, j, $I_i > I_j$ *and* $q_i - I_i < q_j - I_j$.

PROOF: Suppose that

$$I_i > I_j \Rightarrow q_i - I_i \leq q_j - I_j \tag{3.2}$$

for all $1 \leq i, j \leq n$ and for some i, j, $I_i > I_j$ and $q_i - I_i < q_j - I_j$.

Let $(I_1', \ldots, I_n')$ be a new incentive scheme satisfying

$$v_i' + \lambda h(v_i') = v_i + \lambda q_i - \mu \qquad \text{for all } i \tag{3.3}$$

where $v_i = V(I_i)$, $v_i' = V(I_i')$, $\lambda > 0$, and μ is such that

$$\lambda \max_i (q_i - h(v_i)) \geq \mu \geq \lambda \min_i (q_i - h(v_i)). \tag{3.4}$$

If $\lambda = \mu = 0$, then $v_i' = v_i$ solves (3.3). The implicit function theorem therefore implies that (3.3) has a solution as long as λ, μ are small. (Even if h is not differentiable it has left and right hand derivatives.)

It follows from (3.2) and (3.4) that the change to the new incentive scheme has the effect of increasing the lowest I_i's and decreasing the highest ones. For each λ pick μ so that $G(a') + K(a')(\sum_{i=1}^{n}\pi_i(a')v_i') = \max_{a \in A}[G(a) + K(a)(\sum_{i=1}^{n}\pi_i(a) v_i')] = \max_{a \in A}[G(a) + K(a)(\sum_{i=1}^{n}\pi_i(a)v_i)]$. This ensures that the agent's expected

utility remains the same. We now show that the principal's expected profit is higher under the new incentive scheme than under the old, which contradicts the optimality of $(I_1, \ldots, I_n)$.

Substituting (3.1) of Lemma 1 into (3.3) yields:

$$\sum_i \pi_i(a')(q_i - h(v_i')) \geq \sum_i \pi_i(a)(q_i - h(v_i')).$$

If we can show that $\sum \pi_i(a)h(v_i') < \sum \pi_i(a)h(v_i)$, it will follow that

$$\sum \pi_i(a')(q_i - h(v_i')) > \sum \pi_i(a)(q_i - h(v_i)),$$

i.e., the principal is better off.

To see that $\sum \pi_i(a)h(v_i') < \sum \pi_i(a)h(v_i)$, note that

$$\sum \pi_i(a)(h(v_i) - h(v_i')) \geq \sum \pi_i(a)h'(v_i')(v_i - v_i')$$

by the convexity of h (here h' is the right-hand derivative if h is not differentiable). It suffices therefore to show that the latter expression is positive. By (3.3),

$$\sum \pi_i(a)h'(v_i')(v_i - v_i') = \sum \pi_i(a)h'(v_i')(\lambda h(v_i') - \lambda q_i + \mu).$$

Suppose that this is nonpositive for small λ. Divide by λ and let $\lambda \to 0$. Assuming without loss of generality μ/λ converges to $\hat{\mu}$ (we allow $\hat{\mu}$ infinite) and that $h'(v_i')$ converges to $\hat{h}_i'$, and using the fact that $v_i' \to v_i$, we get

$$\sum \pi_i(a)\hat{h}_i'(h(v_i) - q_i + \hat{\mu}) \leq 0. \tag{3.5}$$

However, from the fact that $h'(v_i')$ is nondecreasing in v_i' and $v_i' \to v_i$, $h'(v_i') \to \hat{h}_i'$, it follows that $v_i > v_j \Rightarrow \hat{h}_i' \geq \hat{h}_j'$. Hence by (3.2) $\hat{h}_i'$ and $(h(v_i) - q_i)$ are similarly ordered in the sense of Hardy, Littlewood, and Polya [**5**]; i.e., as one moves up so does the other. Therefore, by Hardy, Littlewood, and Polya [**5**, p. 43], $\hat{h}_i'$ and $(h(v_i) - q_i)$ are positively correlated, i.e.,

$$\sum \pi_i(a)\hat{h}_i'(h(v_i) - q_i + \hat{\mu}) > \Big(\sum \pi_i(a)\hat{h}_i'\Big)\Big(\sum \pi_i(a)(h(v_i) - q_i + \hat{\mu})\Big) \geq 0, \tag{3.6}$$

where the last inequality follows from the fact that (1) $h' \geq 0$; (2) $G(a) + K(a)(\sum \pi_i(a)v_i') \leq G(a') + K(a')(\sum \pi_i(a')v_i') = G(a) + K(a)(\sum \pi_i(a)v_i)$ (since the agent's expected utility stays constant), which implies that

$$\lim_{\lambda \to 0} (1/\lambda) \sum \pi_i(a)(v_i - v_i') \geq 0.$$

(3.6) contradicts (3.5).

This proves that $\sum \pi_i(a)h(v_i') < \sum \pi_i(a)h(v_i)$, which establishes that the principal's expected profit is higher under $(I_1', \ldots, I_n')$. Contradiction. *Q.E.D.*

REMARK 5: Another way of expressing Proposition 4 is that there is no

permutation $i_1, \ldots, i_n$ of the integers $1, \ldots, n$ such that I_{i_k} is nondecreasing in k, and $(q_{i_k} - I_{i_k})$ is nonincreasing in k, with $I_{i_k} < I_{i_{k+1}}$, $(q_{i_k} - I_{i_k}) > (q_{i_{k+1}} - I_{i_{k+1}})$ for some k. Note that there is an interesting contrast between Proposition 4 and results found in the literature on optimal risk sharing in the absence of moral hazard. In this literature (see Borch [**4**]), it is shown that (if the individuals are risk averse) it is optimal for the individuals' returns to be positively related over the *whole* range of outcomes, whereas here we are only able to show that this is true over some range of outcomes.

Proposition 4 may be used to establish the following result about the monotonicity of the optimal incentive scheme.

PROPOSITION 5: *Assume* A1–A3 *and* V *strictly concave. Let* $(I_1, \ldots, I_n)$ *be a second-best optimal incentive scheme. Then* (1) *there exists* $1 \leq i \leq n-1$ *such that* $I_i \leq I_{i+1}$, *with strict inequality unless* $I_1 = I_2 = \cdots = I_n$; (2) *there exists* $1 \leq j \leq n-1$ *such that* $q_j - I_j < q_{j+1} - I_{j+1}$.

PROOF: (1) follows directly from Proposition 4. So does (2) once we rule out the case $q_1 - I_1 = q_2 - I_2 = \cdots = q_n - I_n$. We do this by a similar argument to that used in Proposition 4. Suppose that $\boldsymbol{I}$ is an optimal incentive scheme satisfying

$$q_1 - I_1 = q_2 - I_2 = \cdots = q_n - I_n = k. \tag{3.7}$$

Then $I_1 < I_2 < \cdots < I_n$. Consider the new incentive scheme $\boldsymbol{I}' = (I_1 + \epsilon, I_2 + \epsilon, \ldots, I_{n-1} + \epsilon, I_n - \mu\epsilon)$ where $\epsilon > 0$ and μ is chosen so that $\max_{a \in A}[G(a) + K(a)(\sum \pi_i(a) V(I_i'))] = \max_{A \in A}[G(a) + K(a)(\sum \pi_i(a) V(I_i))]$, i.e. the agent's expected utility is kept constant. We show that the principal's expected profit is higher under $\boldsymbol{I}'$ than under $\boldsymbol{I}$ for small ϵ. Suppose not. Then

$$\sum \pi_i(a')(q_i - I_i') \leq \sum \pi_i(a)(q_i - I_i) = k,$$

where a' (resp. a) is optimal for the agent under $\boldsymbol{I}'$ (resp. $\boldsymbol{I}$). Substituting for $\boldsymbol{I}'$ yields

$$-(1 - \pi_n(a'))\epsilon + \pi_n(a')\mu\epsilon \leq 0.$$

Take limits as $\epsilon \to 0$. Without loss of generality $a' \to \hat{a}$. Hence we have

$$-(1 - \pi_n(\hat{a})) + \pi_n(\hat{a})\mu \leq 0. \tag{3.8}$$

Now since a' is an optimal action for the agent under $\boldsymbol{I}'$, it follows by uppersemicontinuity that $\hat{a}$ is optimal under $\boldsymbol{I}$. Hence we have

$$\begin{aligned} G(\hat{a}) + K(\hat{a})\Big(\sum \pi_i(\hat{a}) V(I_i')\Big) &\leq G(a') + K(a')\Big(\sum \pi_i(a') V(I_i')\Big) \\ &= G(\hat{a}) + K(\hat{a})\Big(\sum \pi_i(\hat{a}) V(I_i)\Big). \end{aligned}$$

Hence $\sum \pi_i(\hat{a})(V(I_i) - V(I_i')) \geq 0$. Using the concavity of V and taking limits as

$\epsilon \to 0$, we get

$$\sum_{i=1}^{n-1} \pi_i(\hat{a}) V'(I_i) - \pi_n(\hat{a}) V'(I_n)\,\mu \leq 0.$$

But since $V'(I_i)$ is decreasing in i, this contradicts (3.8). (If V is not differentiable, V' denotes the right-hand derivative.)

This proves that the principal does better under $\boldsymbol{I}'$ than under $\boldsymbol{I}$. Hence we have ruled out the case $q_1 - I_1 = \cdots = q_n - I_n$. This establishes Proposition 5. *Q.E.D.*

Proposition 5 says that it is not optimal for the agent's marginal reward as a function of income to be negative everywhere or to be greater than or equal to one everywhere.[8] However, the proposition does allow for the possibility that either of these conditions can hold over some interval. To see when this may occur, it is useful to consider in more detail the case where A is a finite set. When A is finite, we can use the Kuhn–Tucker conditions for problem (2.2) to characterize the optimum. If Assumption A3 holds and h is differentiable, these yield:

$$h'(v_i) = \left[\lambda + \sum_{\substack{a_j \in A \\ a_j \neq a^*}} \mu_j\right] K(a^*) - \sum_{\substack{a_j \in A \\ a_j \neq a^*}} \mu_j K(a_j)\left(\frac{\pi_i(a_j)}{\pi_i(a^*)}\right) \qquad \text{for all } i, \tag{3.9}$$

where $\lambda, (\mu_j)$ are nonnegative Lagrange multipliers and $\mu_j > 0$ only if the agent is indifferent between a^* and a_j at the optimum. The following proposition states that $\mu_j > 0$ for at least one action which is less costly than a^*. This implies that at an optimum the agent must be indifferent between at least two actions (unless a^* is the least costly action, i.e. where there is no incentive problem).

PROPOSITION 6: *Assume* A1–A3 *and A finite. Suppose that* (2.2) *has a solution for* $a^* \in A$. *Then if* $C_{FB}(a^*) > \min_{a \in A} C_{FB}(a)$, *this solution will have the property that* $G(a^*) + K(a^*)(\sum_{i=1}^{n} \pi_i(a^*)v_i) = G(a_j) + K(a_j)(\sum_{i=1}^{n} \pi_i(a_j)v_i)$ *for some* $a_j \in A$ *with* $C_{FB}(a_j) < C_{FB}(a^*)$. *Furthermore, if V is strictly concave and differentiable, the Lagrange multiplier* μ_j *will be strictly positive for some* a_j *with* $C_{FB}(a_j) < C_{FB}(a^*)$.

PROOF: Suppose that the agent strictly prefers a^* to all actions less costly than a^* at the solution. Then, since (2.2) is a convex problem, we can drop all the constraints in (2.2) which refer to less costly actions without affecting the

[8] Among other things, Proposition 5 shows that it is not optimal to have $q_1 - I_1 = q_2 - I_2 = \cdots = q_n - I_n$. This result has also been established by Shavell [**20**] under stronger assumptions.

solution. In other words, we can substitute $A' = \{a \in A \mid a$ is at least as costly as $a^*\}$ for A in (2.2) and the solution will not change. But since a^* is now the least costly action, we know from the proof of Proposition 1 that it is optimal to set $I_i = I_j$ for all i, j. However, $I_i = I_j$ is not optimal for the original problem since, under these conditions, the agent will pick an a which minimizes $C_{FB}(a)$, and by assumption $C_{FB}(a^*) > \min_{a \in A} C_{FB}(a)$. Contradiction.

That $\mu_j > 0$ follows from the fact that if all the $\mu_j = 0$, then $h'(v_i)$ is the same for all i, which implies that $I_1 = \cdots = I_n$; however, this means that the agent will choose a cost-minimizing action, contradicting $C_{FB}(a^*) > \min_{a \in A} C_{FB}(a)$. *Q.E.D.*

It should be noted that Proposition 6 depends strongly on the assumption that A is finite.

The simplest case occurs when $\mu_j > 0$ for just one a_j with $C_{FB}(a_j) < C_{FB}(a^*)$ (this will be true in particular if A contains only two actions). In this case, we can rewrite (3.9) as

$$h'(v_i) = (\lambda + \mu_j)K(a^*) - \mu_j K(a_j)\frac{\pi_i(a_j)}{\pi_i(a^*)} . \tag{3.10}$$

We see that what determines v_i, and hence I_i, in this case is the relative likelihood that the outcome $q = q_i$ results from a_j rather than from a^*. In particular, since h convex $\Rightarrow h'$ nondecreasing in v_i, a sufficient condition for the optimal incentive scheme to be nondecreasing everywhere, i.e. $I_1 \leq I_2 \leq \cdots \leq I_n$, is that $\pi_i(a_j)/\pi_i(a^*)$ is nonincreasing in i, i.e. the relative likelihood that $a = a_j$ rather than $a = a^*$ produces the outcome $q = q_i$ is lower the better is the outcome i.

This observation has led some to suggest that the following is a sufficient conditon for the incentive scheme to be nondecreasing.

MONOTONE LIKELIHOOD RATIO CONDITION (MLRC): Assume A3. Then MLRC holds if, given $a, a' \in A$, $C_{FB}(a') \leq C_{FB}(a)$ implies that $\pi_i(a')/\pi_i(a)$ is nonincreasing in i.

It should be noted that the "first-order condition" approach described in the introduction, which is based on the assumption that the agent is indifferent between a and $a + da$ at an optimum, does yield MLRC as a sufficient condition for monotonicity.[9] We now show, however, that, once we take into account the possibility that the agent may be indifferent between several actions at an

[9]See Mirrlees [11] or Holmstrom [7]. Milgrom [9] has shown that MLRC, as stated here, implies the differential version of the monotone likelihood condition which is to be found in Mirrlees [11] or Holmstrom [7].

optimum, i.e. $\mu_j > 0$ for more than one a_j, MLRC does not guarantee monotonicity.

EXAMPLE 1: $A = \{a_1, a_2, a_3\}$, $n = 3$. $\pi(a_1) = (\frac{2}{3}, \frac{1}{4}, \frac{1}{12})$, $\pi(a_2) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $\pi(a_3) = (\frac{1}{12}, \frac{1}{4}, \frac{2}{3})$. Assume additive separability with $G(a_1) = 0$, $G(a_2) = -(\frac{1}{12}\sqrt{2} + \frac{1}{4}\sqrt{7/4})$, $G(a_3) = -\frac{7}{12}\sqrt{7/4}$, $V(I) = (3I)^{1/3}$ (i.e. $h(v) = \frac{1}{3}v^3$), $K(a) \equiv 1$ and $\bar{U} = \frac{1}{4}\sqrt{2} + \frac{1}{12}\sqrt{7/4}$. Note that MLRC is satisfied here.[10]

We compute $C(a_1), C(a_2), C(a_3)$. Obviously, $C(a_1) = C_{FB}(a_1) = \frac{1}{3}(\bar{U} - G(a_1))^3 = 0.033$. To compute $C(a_2)$, we use the first-order conditions (3.9). These are

$$v_1^2 = \lambda - \mu_1 + \tfrac{3}{4}\mu_2,$$

$$v_2^2 = \lambda + \tfrac{1}{4}\mu_1 + \tfrac{1}{4}\mu_2,$$

$$v_3^2 = \lambda + \tfrac{3}{4}\mu_1 - \mu_2,$$

plus the complementary slackness conditions. These equations are solved by setting $\lambda = \frac{5}{4}$, $\mu_1 = 2$, $\mu_2 = 1$. This yields $v_1 = 0$, $v_2 = \sqrt{2}$, $v_3 = \sqrt{7/4}$, and the agent is then indifferent between a_1, a_2, and a_3:

$$\tfrac{2}{3}v_1 + \tfrac{1}{4}v_2 + \tfrac{1}{12}v_3 + G(a_1) = \tfrac{1}{3}v_1 + \tfrac{1}{3}v_2 + \tfrac{1}{3}v_3 + G(a_2)$$

$$= \tfrac{1}{12}v_1 + \tfrac{1}{4}v_2 + \tfrac{2}{3}v_3 + G(a_3) = \bar{U}.$$

Since the first-order conditions are necessary and sufficient, we may conclude that $C(a_2) = \frac{1}{3}(\frac{1}{3}v_1^3 + \frac{1}{3}v_2^3 + \frac{1}{3}v_3^3) = 0.571$.

Note that the incentive scheme which implements a_2, $I_1 = 0$, $I_2 = \frac{1}{3}2^{3/2}$, $I_3 = \frac{1}{3}(\frac{7}{4})^{3/2}$, is not nondecreasing.

Observe that $C(a_3) \geq C_{FB}(a_3) = \frac{1}{3}(\bar{U} - G(a_3))^3 = 0.635 > C(a_2)$. Since $C(a_3) > C(a_2) > C(a_1)$, it is easy to show that we can find $q_1 < q_2 < q_3$ such that $B(a_2) - C(a_2) > \max[B(a_3) - C(a_3), B(a_1) - C(a_1)]$. But this means that it is optimal for the principal to get the agent to pick a_2. Hence the optimal incentive scheme is as described above. It is not nondecreasing despite the satisfaction of MLRC.

The reason that monotonicity breaks down in Example 1 is because, at the optimum, the agent is indifferent between a_2, the action to be implemented, a_1 a less costly action, and a_3 a more costly action. By MLRC $\pi_i(a_1)/\pi_i(a_2), \pi_i(a_2)/\pi_i(a_3)$ are decreasing in i. However, $\mu_1(\pi_i(a_1)/\pi_i(a_2)) + \mu_2(\pi_i(a_3)/\pi_i(a_2))$ need not be monotonic.

This observation suggests that one way to get monotonicity is to strengthen MLRC so that it holds for weighted combinations of actions as well as for the

[10]The function V violates (2) of Assumption A1, but this is unimportant for the example.

basic actions themselves. In particular, suppose that

(3.11) given any finite subset $\{a_1, \ldots, a_m\}$ of A, $a \in A$, and nonnegative weights $w_1, \ldots, w_m$ summing to 1, it is the case that $\left(\sum_{j=1}^{m} w_j \pi_i(a_j)/\pi_i(a) \right)$ is either nondecreasing in i or nonincreasing in i.

Then, by the first-order conditions (3.9),

$$h'(v_i) = \left[\lambda + \sum_{\substack{a_j \in A \\ a_j \neq a^*}} \mu_j \right] K(a^*) - \left[\sum_{\substack{a_j \in A \\ a_j \neq a^*}} \mu_j K(a_j) \right] \left[\sum_{\substack{a_j \in A \\ a_j \neq a^*}} w_j \left(\frac{\pi_i(a_j)}{\pi_i(a^*)} \right) \right], \tag{3.12}$$

where

$$w_j = \mu_j K(a_j) \Big/ \sum_{\substack{a_h \in A \\ a_h \neq a^*}} \mu_h K(a_h).$$

But, by (3.11), the right-hand side (RHS) of (3.12) is monotonic. Hence, the v_i's are either monotonically nondecreasing or nonincreasing. By Proposition 5, however, they cannot be nonincreasing; hence they are nondecreasing.

Unfortunately, (3.11) turns out to be a very strong condition. In fact, it is equivalent to the following spanning condition.

SPANNING CONDITION (SC): There exists $\hat{\pi}, \hat{\pi}' \in S$ such that (1) for each $a \in A$, $\pi(a) = \lambda(a)\hat{\pi} + (1 - \lambda(a))\hat{\pi}'$ for some $0 \leq \lambda(a) \leq 1$; (2) $\hat{\pi}_i/\hat{\pi}_i'$ is nonincreasing in i.

That SC implies (3.11) is easy to see. We are grateful to Jim Mirrlees for pointing out and proving the converse.[11]

PROPOSITION 7: *Assume* A1–A3, *V strictly concave and differentiable. Suppose that SC holds. Then a second-best optimal incentive scheme satisfies* $I_1 \leq I_2 \leq \cdots \leq I_n$.

PROOF: If A is finite, the argument following (3.12) establishes the result. To establish the result for the case A infinite, let $\hat{a} \in A$ be a second-best optimal

[11] To prove the converse, define $a \lesssim a'$ if $\pi_i(a')/\pi_i(a)$ is nondecreasing in i. (3.11) implies that $\lesssim$ is a complete pre-ordering on A. Furthermore, $\lesssim$ is continuous. Since A is compact, there exist $\underline{a}, \bar{a} \in A$ such that $\underline{a} \lesssim a \lesssim \bar{a}$ for all $a \in A$. Given $a \in A$, consider $\lambda(\pi_i(\bar{a})/\pi_i(a)) + (1 - \lambda)(\pi_i(\underline{a})/\pi_i(a))$. When $\lambda = 1$, this is nondecreasing in i, and when $\lambda = 0$, it is nonincreasing in i. Furthermore, (3.11) implies that it is monotonic in i for all $0 < \lambda < 1$. It follows by continuity that it is independent of i for some $0 < \lambda < 1$.

action and let $\boldsymbol{I}$ be the second-best optimal incentive scheme which implements it. By Remark 2 of Section 2, $\boldsymbol{I}$ is unique. Let A_r be a finite subset of A containing $\hat{a}$ such that the Euclidean distance between A_r and A is less than $(1/r)$. Let $\boldsymbol{I}_r$ be the second-best optimal incentive scheme which implements $\hat{a}$ when the agent is restricted to choosing from A_r. From Proposition 7 for the finite A case, we know that $\boldsymbol{I}_r$ is nondecreasing. Take limits as $r \to \infty$. It is straightforward to show that $I_r \to \boldsymbol{I}$. It follows that $\boldsymbol{I}$ is nondecreasing. *Q.E.D.*

An alternative sufficient condition for monotonicity may be found in the work of Mirrlees [**12**], who establishes a similar result to Proposition 8 below. For each $a \in A$, let $F(a) = (\pi_1(a), \pi_1(a) + \pi_2(a), \ldots, \pi_1(a) + \cdots + \pi_n(a))$. In the following proposition, the notation $F(a) \geq F'(a)$ is used to mean $F_i(a) \geq F_i'(a)$ for all $i = 1, \ldots, n$.

CONCAVITY OF DISTRIBUTION FUNCTION CONDITION (CDFC): CDFC holds if $a, a', a'' \in A$, and

$$\left(\frac{\bar{U} - G(a)}{K(a)}\right) = \lambda\left(\frac{\bar{U} - G(a')}{K(a')}\right) + (1-\lambda)\left(\frac{\bar{U} - G(a'')}{K(a'')}\right), \qquad 0 \leq \lambda \leq 1,$$

imply that $F(a) \leq \lambda F(a') + (1-\lambda)F(a'')$.

PROPOSITION 8: *Assume* A1–A3, *V strictly concave and differentiable. Assume also that U is additively or multiplicatively separable, i.e., either* $G(a) \equiv 0$ *or* $K(a) \equiv$ *constant. Suppose that MLRC and CDFC hold. Then a second-best optimal incentive scheme* $(I_1, \ldots, I_n)$ *satisfies* $I_1 \leq I_2 \leq \cdots \leq I_n$.

PROOF: Assume first that A is finite. Let a^* maximize $B(a) - C(a)$. Let $A' = \{a \in A \mid C_{FB}(a) \leq C_{FB}(a^*)\}$. Consider the cost minimizing way of getting the agent to pick a^* given that he can choose only from A'. It is clear from (3.9) that, since $\pi_i(a_j)/\pi_i(a^*)$ is nonincreasing in i by MLRC, the incentive scheme $(I_1, \ldots, I_n)$ is nondecreasing. We will be home if we can show that $(I_1, \ldots, I_n)$ is optimal when A' is replaced by A. Since adding actions cannot reduce the cost of implementing a^*, all we have to do is to show that $(I_1, \ldots, I_n)$ continues to implement a^*, i.e. there does not exist a'', $C_{FB}(a'') > C_{FB}(a^*)$, such that

$$(3.13) \qquad G(a'') + K(a'')\left(\sum \pi_i(a'')v_i\right) > G(a^*) + K(a^*)\left(\sum \pi_i(a^*)v_i\right).$$

However, we know from Propositions 2 and 6 that

$$(3.14) \qquad G(a^*) + K(a^*)\left(\sum \pi_i(a^*)v_i\right) = G(a') + K(a')\left(\sum \pi_i(a')v_i\right) = \bar{U}$$

for some a' with $C_{FB}(a') < C_{FB}(a^*)$. Writing

$$\frac{\bar{U} - G(a^*)}{K(a^*)} = \lambda\left(\frac{\bar{U} - G(a'')}{K(a'')}\right) + (1-\lambda)\left(\frac{\bar{U} - G(a')}{K(a')}\right)$$

and using CDFC and the fact that $v_1 \leq v_2 \leq \cdots \leq v_n$, we get

$$\sum \pi_i(a^*)v_i - \left(\frac{\bar{U} - G(a^*)}{K(a^*)}\right)$$

$$\geq \lambda \sum \pi_i(a'')v_i + (1-\lambda)\left(\sum \pi_i(a')v_i\right) - \left(\frac{\bar{U} - G(a^*)}{K(a^*)}\right)$$

$$= \lambda\left[\sum \pi_i(a'')v_i - \left(\frac{\bar{U} - G(a'')}{K(a'')}\right)\right]$$

$$+ (1-\lambda)\left[\sum \pi_i(a')v_i - \left(\frac{\bar{U} - G(a')}{K(a')}\right)\right].$$

But this contradicts (3.13) and (3.14).

To prove the result for A finite, one again proceeds by way of finite approximation. *Q.E.D.*

To understand CDFC, consider, for each $a \in A$, $V(C_{FB}(a)) = ((\bar{U} - G(a))/K(a))$. In utility terms $V(C_{FB}(a))$ is a measure of the first-best cost of getting the agent to pick a. CDFC says that if a is a convex combination of a' and a'' in terms of this measure of cost then the distribution function of outcomes corresponding to a dominates in the sense of first degree stochastic dominance the corresponding convex combination of the distribution functions corresponding to a' and a''. It is worth noting that under the assumption of additive or multiplicative separability in Proposition 8, the λ in the CDFC definition is independent of $\bar{U}$.

So far we have considered only the monotonicity of the optimal incentive scheme. One would also like to know when the optimal incentive scheme is *progressive*, i.e. $(I_{i+1} - I_i)/(q_{i+1} - q_i)$ is nonincreasing in i, or *regressive*, i.e. $(I_{i+1} - I_i)/(q_{i+1} - q_i)$ is nondecreasing in i. To get results about this, one needs considerably stronger assumptions, as the following proposition indicates.

PROPOSITION 9: *Assume* A1–A3, *V strictly concave and differentiable. Assume also that U is additively or multiplicatively separable, i.e., either $G(a) \equiv 0$ or $K(a) \equiv$ constant. Suppose that MLRC and CDFC hold and that $(q_{i+1} - q_i)$ is independent of i, $1 \leq i \leq n-1$. Then a second-best optimal incentive scheme will be regressive (resp. progressive) if*

(3.15) *$(1/V'(I))$ is concave (resp. convex) in I and $a, a' \in A$,*

$C_{FB}(a') < C_{FB}(a)$, implies that $(\pi_{i+1}(a')/\pi_{i+1}(a)) - (\pi_i(a')/\pi_i(a))$

is nonincreasing (resp. nondecreasing) in i.

PROOF: Assume first that A is finite. Let a^* be a second-best optimal action. Let a' maximize $C_{FB}(a)$ subject to $C_{FB}(a) < C_{FB}(a^*)$, i.e. a' is the next most costly action after a^*. Consider the cost minimizing way of implementing a^* given that a' is the only other action that the agent can choose. Using the same concavity argument as in the proof of Proposition 8, we can show that the resulting incentive scheme $(I_1, \ldots, I_n)$ also implements a^* when the agent can choose from all of A. Hence $(I_1, \ldots, I_n)$ is an optimal incentive scheme.

By (3.10),

$$\frac{1}{V'(I_i)} = h'(v_i) = (\lambda + \mu)K(a^*) - \mu K(a')\frac{\pi_i(a')}{\pi_i(a^*)}$$

and so

$$\frac{1}{V'(I_{i+1})} - \frac{1}{V'(I_i)} = -\mu K(a')\left(\frac{\pi_{i+1}(a')}{\pi_{i+1}(a^*)} - \frac{\pi_i(a')}{\pi_i(a^*)}\right).$$

(3.15) now follows immediately. To prove the result for the A infinite case, one again proceeds by way of a finite approximation. *Q.E.D.*

Note that $1/V'$ is linear if $V = \log I$; is concave if $V = -e^{-\alpha I}$, $\alpha > 0$, or $V = I^{\alpha}$, $0 < \alpha < 1$; is convex if $V = -I^{-\alpha}$, $\alpha > 1$.

It should also be noted that Mirrlees [**12**] has shown that if CDFC holds, the "first-order condition" approach referred to in the introduction is valid. Thus Propositions 8 and 9 can also be proved by appealing to the characterization of an optimal incentive scheme to be found in much of the literature (see, e.g., Holmstrom [7] and Mirrlees [**11**]).

Let us summarize the results of this section. We have shown that an optimal incentive scheme will not be declining everywhere, but that only under quite strong assumptions (SC or MLRC plus concavity) will it be nondecreasing everywhere. We have also shown that it is not optimal for the agent's marginal remuneration for an extra pound of profit to exceed one everywhere, although it may exceed one sometimes. Finally, we have obtained sufficient conditions for the incentive scheme to be progressive or regressive.

The conclusion that only under strong assumptions will the optimal incentive scheme be monotonic may seem disappointing at first sight. One feels that monotonicity is a minimal requirement. This may not be the right reaction, however. There are many interesting situations where it is clear that the optimal scheme will not be monotonic. We have described one example in the introduction. Another example is the following. Suppose that actions are two dimensional, with one dimension referring to how hard the agent works and the other dimension to how cautious he is—greater caution might lead to a lower variance of profit but also to a lower mean. The optimal action for the principal might involve the agent working fairly hard and also not being too cautious. The best

way to implement this may be to pay the agent high amounts for both very good outcomes (to encourage high effort) and very bad outcomes (to discourage excessive caution). This example seems far from pathological. In fact, one might argue that a number of real world incentive schemes operate in this way. In view of examples like this, the difficulty of finding general conditions guaranteeing monotonicity may become less surprising.[12]

In the next section, we show that considerably stronger results than those of this section can be proved for the case $n = 2$. We also provide a simple algorithm for computing optimal incentive schemes when $n = 2$.

4. THE CASE OF TWO OUTCOMES

When $n = 2$, we will refer to q_1 as the "bad" outcome and $q_2 > q_1$ as the "good" outcome. In this case, the agent's incentive scheme can be represented simply by a fixed payment w and a share of profits, s, where $w + sq_1 = I_1$, $w + sq_2 = I_2$, i.e., $s = (I_2 - I_1)/(q_2 - q_1)$. Proposition 5 of the last section shows that it is not optimal for I_i to be everywhere declining in q_i. When $n = 2$, this means that $s \geq 0$.[13] Similarly the proposition implies that $s < 1$ when $n = 2$. This has a number of interesting implications.

DEFINITION: Let $n = 2$. We say that $a \in A$ is *efficient* if there does not exist $a' \in A$ satisfying $C_{FB}(a') \leq C_{FB}(a)$ and $\pi_2(a') \geq \pi_2(a)$, with at least one strict inequality.

In other words, an action is efficient if the probability of a good outcome can only be increased by incurring greater cost.

PROPOSITION 10: *Assume* A1–A3 *and* V *strictly concave. Let* $n = 2$. *Then every second-best optimal action is efficient.*

PROOF: Let a be a second-best optimal action. Then a maximizes $G(a) + K(a)[\pi_1(a)v_1 + \pi_1(a)v_2]$. Suppose $C_{FB}(a') \leq C_{FB}(a)$ and $\pi_2(a') \geq \pi_2(a)$, with at least one strict inequality. Then, by the definition of C_{FB},

$$G(a) + K(a)V(C_{FB}(a)) = \bar{U} = G(a') + K(a')V(C_{FB}(a'))$$
$$\leq G(a') + K(a')V(C_{FB}(a)),$$

[12] There are some cases where monotonicity may be a *constraint* on the optimal incentive scheme. An example is where the agent can always make a better outcome look like a worse outcome by reducing the firm's profits after the outcome has occurred. This case can be analyzed by adding the (linear) constraints $v_1 \leq v_2 \leq \cdots \leq v_n$ to the problem (2.2).

[13] Shavell [**19**] also proves that $s \geq 0$ when $n = 2$, but under stronger assumptions.

since $C_{FB}(a') \leq C_{FB}(a)$. Hence, by Assumption A1(4), $G(a) + K(a)v \leq G(a') + K(a')v$ for all $v \in \mathfrak{V}$. Therefore using the fact that $v_1 \leq v_2$ since $s \geq 0$, and the fact that $\pi_2(a') \geq \pi_2(a)$, we have

$$\begin{aligned} & G(a) + K(a)\left[\pi_1(a)v_1 + \pi_2(a)v_2\right] \\ & \quad \leq G(a') + K(a')\left[\pi_1(a)v_1 + \pi_2(a)v_2\right] \\ & \quad \leq G(a') + K(a')\left[\pi_1(a')v_1 + \pi_2(a')v_2\right] \end{aligned}$$

with at least one strict inequality unless $C_{FB}(a) = C_{FB}(a')$ and $v_1 = v_2$. This contradicts the optimality of a unless $C_{FB}(a) = C_{FB}(a')$ and $v_1 = v_2$. However, in this case, the agent is indifferent between a and a', while the principal prefers a', again contradicting the optimality of a. *Q.E.D.*

We may use Proposition 10 to prove that when $n = 2$ it will never pay the principal to offer the agent an expected utility in excess of $\overline{U}$ (recall that when $n > 2$ this is only generally true when $U(a, I)$ is additively or multiplicatively separable—see Proposition 2).

PROPOSITION 11: *Assume* A1–A3 *and V strictly concave. Let* $n = 2$. *Let* $\hat{a}$ *be a second-best optimal action and* $\hat{\mathbf{I}}$ *a second-best optimal incentive scheme which implements* $\hat{a}$. *Then* $\sum_{i=1}^{n} \pi_i(\hat{a}) U(\hat{a}, \hat{I}_i) = \overline{U}$.

PROOF: Suppose not, i.e., $\sum_{i=1}^{n} \pi_i(\hat{a}) U(\hat{a}, \hat{I}_i) > \overline{U}$. Consider a new incentive scheme $(I_1, I_2) = (\hat{I}_1 - \epsilon, \hat{I}_2)$ where $\epsilon > 0$ is small. Let a be an optimal action for the agent under the new scheme, i.e., a maximizes $G(a) + K(a)[\pi_1(a)V(\hat{I}_1 - \epsilon) + \pi_2(a)V(\hat{I}_2)]$. Then,

$$\begin{aligned} \pi_1(a)(q_1 - I_1 + \epsilon) + \pi_2(a)(q_2 - I_2) &> \pi_1(a)(q_1 - I_1) + \pi_2(a)(q_2 - I_2) \\ &\geq \pi_1(\hat{a})(q_1 - \hat{I}_1) + \pi_2(\hat{a})(q_2 - \hat{I}_2) \end{aligned}$$

as long as $\pi_2(\hat{a}) \leq \pi_2(a)$ (since $0 \leq s < 1$). Thus, if we can show that $\pi_2(\hat{a}) \leq \pi_2(a)$, we will have contradicted the optimality of $(\hat{I}_1, \hat{I}_2)$, since the principal's profits will be higher under (I_1, I_2) than under $(\hat{I}_1, \hat{I}_2)$.

Suppose $\pi_2(\hat{a}) > \pi_2(a)$. Now the same argument as in Proposition 10 shows that a is efficient. Thus we must have $C_{FB}(\hat{a}) > C_{FB}(a)$. Hence $G(a) + K(a)V(C_{FB}(a)) = \overline{U} = G(\hat{a}) + K(\hat{a})V(C_{FB}(\hat{a})) > G(\hat{a}) + K(\hat{a})V(C_{FB}(a))$, and so, by Assumption A1(4),

$$G(a) + K(a)v > G(\hat{a}) + K(\hat{a})v \tag{4.1}$$

for all $v \in \mathfrak{V} \equiv \{V(I) \mid I \in \mathcal{I}\}$. Since $\mathfrak{V}$ contains arbitrarily large negative num-

bers, we may conclude from (4.1) that $K(a) \leq K(\hat{a})$. Now by revealed preference,

$$(4.2)\qquad G(a) + K(a)\left[\pi_1(a)V(\hat{I}_1) + \pi_2(a)V(\hat{I}_2)\right] \leq G(\hat{a}) + K(\hat{a})\left[\pi_1(\hat{a})V(\hat{I}_1) + \pi_2(\hat{a})V(\hat{I}_2)\right],$$

$$(4.3)\qquad G(a) + K(a)\left[\pi_1(a)V(\hat{I}_1 - \epsilon) + \pi_2(a)V(\hat{I}_2)\right] \geq G(\hat{a}) + K(\hat{a})\left[\pi_1(\hat{a})V(\hat{I}_1 - \epsilon) + \pi_2(\hat{a})V(\hat{I}_2)\right].$$

Subtracting (4.3) from (4.2) yields $K(a)\pi_1(a) \leq K(\hat{a})\pi_1(a)$. Hence, since $\pi_2(\hat{a}) > \pi_2(a)$ by assumption, $K(a) < K(\hat{a})$. However, rewriting (4.2), we obtain

$$G(a) + K(a)\bar{v} + K(a)\left[\pi_1(a)(V(\hat{I}_1) - \bar{v}) + \pi_2(a)(V(\hat{I}_2) - \bar{v})\right] \leq G(\hat{a}) + K(\hat{a})\bar{v} + K(\hat{a})\left[\pi_1(\hat{a})(V(\hat{I}_1) - \bar{v}) + \pi_2(\hat{a})(V(\hat{I}_2) - \bar{v})\right]$$

where $\bar{v} = \sup \mathfrak{A}$. (Note that $\bar{v} < \infty$, for $\bar{v} = \infty$ and $K(a) < K(\hat{a})$ violate (4.1).) Setting $v = \bar{v}$ in (4.1), we may conclude that

$$K(a)\pi_1(a)(V(\hat{I}_1) - \bar{v}) + K(a)\pi_2(a)(V(\hat{I}_2) - \bar{v}) \leq K(\hat{a})\pi_1(\hat{a})(V(\hat{I}_1) - \bar{v}) + K(\hat{a})\pi_2(\hat{a})(V(\hat{I}_2) - \bar{v}).$$

But this is impossible since $K(a)\pi_1(a) \leq K(\hat{a})\pi_1(\hat{a})$, $K(a) < K(\hat{a})$, $\pi_2(a) < \pi_2(\hat{a})$, $V(\hat{I}_1) - \bar{v} < 0$, $V(\hat{I}_2) - \bar{v} < 0$. We have thus shown that $\pi_2(a) \geq \pi_2(\hat{a})$, which contradicts the optimality of $(\hat{I}_1, \hat{I}_2)$. *Q.E.D.*

Proposition 11 tells us that the agent's fixed payment w is determined once s is. In particular, w will be the unique solution of

$$\max_{a \in A}\left[G(a) + K(a)(\pi_1(a)V(w + sq_1) + \pi_2(a)V(w + sq_2))\right] = \bar{U}.$$

We have shown that one implication of Proposition 5 for the case $n = 2$ is that every second-best optimal action is efficient. We consider now a second implication. Suppose that we start off in the situation where the agent has access to a set of actions A, and now some additional actions become available, so that the new action set is $A' \supset A$. Then, if the new actions are all higher cost actions for the agent than those in A—in the sense that their C_{FB}'s are higher—the principal cannot be made worse off by such a change.

PROPOSITION 12: *Assume* A1 *and* A2. *Let* $n = 2$. *Suppose that* $A' \supset A$ *and that* $a \in A$, $a' \in A' \backslash A \Rightarrow C_{FB}(a') \geq C_{FB}(a)$. *Assume that* A3 *holds for both* A *and* A'. *Then* $\max_{a \in A'}[B(a) - C'(a)] \geq \max_{a \in A}[B(a) - C(a)]$, *where* C' *is the second-best cost function under* A'.

PROOF: Suppose (I_1, I_2) is an optimal second-best incentive scheme when the action set is A. Let the principal keep this incentive scheme when the new actions

$A' \backslash A$ are added. The only way that the principal can be made worse off is if the agent now switches from $a \in A$ to $a' \in A' \backslash A$. But a' must then provide higher utility for the agent, i.e., $G(a') + K(a')[\pi_1(a')v_1 + \pi_2(a')v_2] > G(a) + K(a)[\pi_1(a)v_1 + \pi_2(a)v_2]$. Since $C_{FB}(a') \geq C_{FB}(a)$, however, $G(a') + K(a')v \leq G(a) + K(a)v$ for all $v \in \mathcal{V}$ (by Assumption A1(4)). Hence $\pi_1(a')v_1 + \pi_2(a')v_2 > \pi(a)v_1 + \pi_2(a)v_2$, which implies, since $v_2 \geq v_1$ by Proposition 5, that $\pi_2(a') > \pi_2(a)$. But it follows that the principal's expected profits $\pi_1(q_1 - I_1) + \pi_2(q_2 - I_2)$ rise when the agent moves from a to a' since, again by Proposition 5, $s < 1$, i.e. $q_2 - I_2 > q_1 - I_1$. *Q.E.D.*

As a final implication of Proposition 5, when $n = 2$, consider a manager-entrepreneur who initially owns 100 per cent of a firm, i.e. $\tilde{w} = 0$, $\tilde{s} = 1$. In the absence of any risk-sharing possibilities the manager will choose a to maximize $\pi_1(a)U(a, q_1) + \pi_2(a)U(a, q_2)$. Let $\tilde{a}$ be a solution to this. Clearly $\tilde{a}$ is efficient. Now suppose a risk neutral principal appears with whom the manager can share risks. We know from Proposition 5 that at the new optimum $s < 1 = \tilde{s}$. Therefore, by Lemma 1 and Proposition 11, $\pi_2(a^*) \leq \pi_2(\tilde{a})$. In addition, $C_{FB}(a^*) \leq C_{FB}(\tilde{a})$ by Proposition 10. Thus, the existence of risk-sharing possibilities leads the agent to choose a less costly action with a lower probability of a good outcome.

We may use Propositions 10–12 to develop a method for computing a second-best optimal incentive scheme when $n = 2$. Consider the case where A is finite. Recall that Proposition 6 states that, in this case, the agent will be indifferent between a^* and some less costly action. This fact makes the computation of an optimal incentive scheme fairly straightforward. We know from Proposition 10 that it is never optimal to get the agent to choose an inefficient action. Hence we can assume without loss of generality that $C_{FB}(a_1) < C_{FB}(a_2) < \cdots < C_{FB}(a_m)$ and $\pi_2(a_1) < \pi_2(a_2) < \cdots < \pi_2(a_m)$. The computation of $C(a_1)$ is easy: by Remark 1 of Section 2 it is just $C_{FB}(a_1)$. To compute $C(a_k)$, $k > 1$, we use Propositions 6 and 11. For each action $a_j, j < k$, find I_1, I_2 so that the agent is indifferent between a_k and a_j and the agent's expected utility is $\bar{U}$. This means solving

$$\begin{aligned} G(a_k) + K(a_k)(\pi_1(a_k)v_1 + \pi_2(a_k)v_2) &= \bar{U}, \\ G(a_j) + K(a_j)(\pi_1(a_j)v_1 + \pi_2(a_j)v_2) &= \bar{U}, \end{aligned} \tag{4.4}$$

which yields

$$\begin{aligned} v_1 &= \frac{\pi_2(a_j)\big(\big(\bar{U} - G(a_k)\big)/K(a_k)\big) - \pi_2(a_k)\big(\big(\bar{U} - G(a_j)\big)/K(a_j)\big)}{\pi_1(a_k) - \pi_1(a_j)}, \\ v_2 &= \frac{\pi_1(a_j)\big(\big(\bar{U} - G(a_k)\big)/K(a_k)\big) - \pi_1(a_k)\big(\big(\bar{U} - G(a_j)\big)/K(a_j)\big)}{\pi_2(a_k) - \pi_2(a_j)}. \end{aligned} \tag{4.5}$$

We then set $I_1 = h(v_1)$, $I_2 = h(v_2)$. Note that $v_1 < v_2$ in (4.5) so that $I_1 < I_2$.

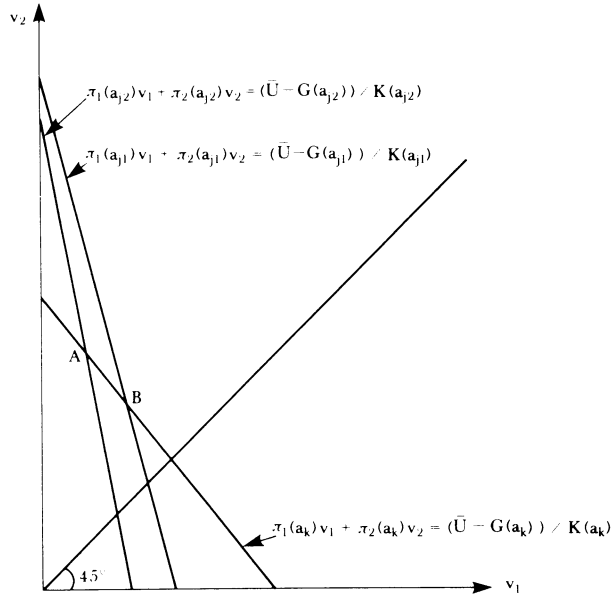


FIGURE 2.

Doing this for each $j = 1, \ldots, k-1$ yields $(k-1)$ different (v_1, v_2) (and (I_1, I_2)) pairs, each with $v_1 < v_2$. This is illustrated in Figure 2 for the case $k = 3$, where the (v_1, v_2) pairs are at A, B. We know from Proposition 6 that one of these pairs is the solution to (2.2). In fact, the solution must occur at the (v_1, v_2) pair with the *smallest* v_1 (and hence, by (4.4), with the largest v_2)—denote this pair by $(\hat{v}_1, \hat{v}_2)$. To see this, suppose that the agent is indifferent between a_k and a_j under $(\hat{v}_1, \hat{v}_2)$. Consider the expression

$$(4.6) \qquad \pi_1(a_k)v_1 + \pi_2(a_k)v_2 - \pi_1(a_j)v_1 - \pi_2(a_j)v_2$$
$$= (\pi_1(a_k) - \pi_1(a_j))v_1 + (\pi_2(a_k) - \pi_2(a_j))v_2 .$$

When $v_1 = \hat{v}_1$, $v_2 = \hat{v}_2$, this expression equals $[(\bar{U} - G(a_k))/K(a_k)] - [(\bar{U} - G(a_j))/K(a_j)]$. Suppose now that $v_1 > \hat{v}_1$, $v_2 < \hat{v}_2$. Then (4.6) falls since $\pi_1(a_k) < \pi_1(a_j)$. Hence the agent now prefers a_j to a_k and so a_k is not implemented.

In Figure 2, the solution is at A. (The solution could not be at B since it is clear from the diagram that, at B, a_{j2} gives the agent an expected utility greater than $\bar{U}$, i.e. a_k is not implemented at B.) Note that it is possible that the $(\hat{v}_1, \hat{v}_2)$ picked in this way does not lie in $\mathfrak{V} \times \mathfrak{V}$; i.e., $h(\hat{v}_1)$ or $h(\hat{v}_2)$ may be undefined. In this case, the constraint set of (2.2) is empty and so $C(a_k) = \infty$. If $(\hat{v}_1, \hat{v}_2) \in \mathfrak{V} \times \mathfrak{V}$, then the principal's minimum expected cost of getting the agent to pick a_k is $C(a_k) = \pi_1(a_k)h(\hat{v}_1) + \pi_1(a_k)h(\hat{v}_2)$. The expected net benefits of implementing a_k are $B(a_k) - C(a_k)$. This procedure must be undergone for each a_k,

$k = 1, \ldots, m$. Finally, the overall optimum is determined by selecting the a which maximizes $B(a_k) - C(a_k)$.

REMARK 6: In computing the cost of implementing a_k, we have ignored actions which are more costly for the agent than a_k. This means that the cost function which we have computed is not the true cost function $C(a)$ but a modified cost function $\tilde{C}(a)$. Clearly, $\tilde{C}(a) \leq C(a)$ for each a since more actions can only make implementation more difficult. On the other hand, Proposition 12 tells us that $\max_{a \in A}[B(a) - \tilde{C}(a)] \leq \max_{a \in A}[B(a) - C(a)]$. Combining these yields $\max_{a \in A}[B(a) - C(a)] = \max_{a \in A}[B(a) - \tilde{C}(a)]$, which means that we are justified in working with $\tilde{C}(a)$ instead of $C(a)$.

Another case where computation is quite simple is when A is infinite and $\{C_{FB}(a) \mid a \in A\}$ is an interval $[\underline{c}, \bar{c}]$ of the real line. For reasons of space, we do not cover this case.

Unfortunately, the computational techniques presented above do not appear to generalize in a useful way to the case $n > 2$. In order to compute an optimum when $n > 2$, in the finite action case, it seems that we must, for each $a \in A$, solve the convex problem in (2.2) and then, by inspection, find the $a \in A$ which maximizes $B(a) - C(a)$. If A is infinite, one takes a finite approximation. These steps can be carried out on a computer, although the amount of computer time involved when the number of elements of A is large may be considerable.

One case where a considerable simplification can be achieved when $n > 2$ is where MLRC and CDFC hold. Then the solution to (2.2) has the property that (1) if A is finite, the agent is indifferent only between a^*, the action the principal wants to implement, and a', where a' maximizes $C_{FB}(a)$ subject to $C_{FB}(a) < C_{FB}(a^*)$, i.e. a' is the next most costly action after a^* (see the proof of Proposition 8); (2) if A is convex, then a^* is the unique maximizer of $G(a) + K(a)(\sum \pi_i(a) V(I_i))$, and $[d(G(a^*) + K(a^*)(\sum \pi_i(a^*) V(I_i)))/da] = 0$ is a necessary and sufficient condition for the agent to pick a^*. In the latter case, Mirrlees [12] has shown that the first-order condition approach referred to in the introduction is valid.

One may ask also whether Propositions 10 and 12 hold in the case $n > 2$. The answer is no (but see Remark 7 below). Second-best optimal actions may be inefficient; i.e., there may exist lower cost actions which dominate the optimal action in the sense of first degree stochastic dominance.[14] Also the addition of actions costlier than the second-best optimal action may make the principal worse off (in Example 1, the principal's expected profits increase if action a_3

[14] Let $A = \{a_1, a_2, a_3\}$, $n = 3$. Assume $C_{FB}(a_1) < C_{FB}(a_2) < C_{FB}(a_3)$, and that $\pi(a_1) = (3/4, 1/8, 1/8)$, $\pi(a_2) = (1/3, 1/3, 1/3)$, $\pi(a_3) = (1/2, 1/2, 0)$ (Assumption A3 is violated, but this is unimportant.) Then $C(a_1) = C_{FB}(a_1)$ since a_1 is the least cost action, and $C(a_3) = C_{FB}(a_3)$ since a_3 can be implemented by setting $I_1 = I_2$, $I_3 = -\infty$. However, $C(a_2) > C_{FB}(a_2)$ and, in fact, if the agent is very risk averse, $C(a_2)$ will be so big that it is profitable for the principal to implement a_3 rather than a_2 (the effect of risk aversion on $C(a)$ is discussed in Section 5). This is in spite of the fact that a_3 is inefficient relative to a_2.

becomes unavailable to the agent). Finally, as Shavell [19] has noted the agent may choose a higher cost action when there are opportunities to share risks with a principal than in the absence of these opportunities.

REMARK 7: It is interesting to note that it is possible to extend all the results of the $n = 2$ case to the $n > 2$ case when the spanning condition (SC) holds. This is because when SC holds, both the principal and the agent are essentially choosing between lotteries of the probability vectors $\hat{\pi}$ and $\hat{\pi}'$.

In particular, let $I_1(v_1) = \min_{\{I_i\}} \sum_{i=1}^n \hat{\pi}_i I_i$ subject to $\sum_{i=1}^n \hat{\pi}_i V(I_i) \geq v_1$; $I_2(v_2) = \min_{\{I_i\}} \sum_{i=1}^n \hat{\pi}_i' I_i$ subject to $\sum_{i=1}^n \hat{\pi}_i' V(I_i) \geq v_2$. Now consider the principal's minimum cost problem as: for each a^*, choose v_1 and v_2 to minimize $\lambda(a^*) I_1(v_1) + (1 - \lambda(a^*))I_2(v_2)$ subject to (1) $G(a^*) + [\lambda(a^*)v_1 + (1 - \lambda(a^*))v_2]K(a^*) \geq G(a) + [\lambda(a)v_1 + (1 - \lambda(a))v_2]K(a)$ for all $a \in A$; (2) $G(a^*) + [\lambda(a^*)v_1 + (1 - \lambda(a^*))v_2]K(a^*) \geq \bar{U}$. Then the principal's problem looks exactly the same as in the $n = 2$ case. Note that from stochastic dominance (i.e. part (2) of the SC condition) $\sum_{i=1}^n \pi_i q_i \leq \sum_{i=1}^n \pi_i' q_i$, so "state 2" is the good state. We are grateful to Bengt Holmstrom for alerting us to the fact that all of the results for the $n = 2$ case hold when $n > 2$ and the Spanning Condition is satisfied.

5. WHAT DETERMINES HOW SERIOUS THE INCENTIVE PROBLEM IS?

In previous sections, we have studied the properties of an optimal incentive scheme. We turn now to a consideration of the factors which determine the magnitude of L, the loss to the principal from being unable to observe the agent's action.

One feels intuitively that the worse is the quality of the information about the agent's action that the principal obtains from observing any outcome, the more serious will be the incentive problem. This idea can be formalized as follows. Suppose that we start with an incentive problem in which the agent's action set is A, his utility function is U, his reservation utility is $\bar{U}$, the probability function is π, and the vector of outputs is $q = (q_1, \ldots, q_n)$. We denote this incentive problem by $(A, U, \bar{U}, \pi, q)$. Consider the new incentive problem $(A, U, \bar{U}, \pi', q')$ where $\pi'(a) = R\pi(a)$ for all $a \in A$ and R is an $(n \times n)$ stochastic matrix (here $\pi(a)$, $\pi'(a)$ are n dimensional column vectors and the columns of R sum to one). Below we show that $C'(a) \geq C(a)$ for all $a \in A$, where unprimed variables refer to the original incentive problem and primed variables to the new incentive problem.

The transformation from $\pi(a)$ to $R\pi(a)$ corresponds to a decrease in informativeness in the sense of Blackwell (see, e.g., Blackwell and Girshick [3]).[15] That is, if we think of the actions $a \in A$ as being parameters with respect to which we

[15] The possibility of using Blackwell's notion of informativeness to characterize the seriousness of an incentive problem was suggested by Holmstrom [7].

have a prior probability distribution, then an experimenter who makes deductions about a from observing $q_1, \ldots, q_n$ would prefer to face the function π than the function $R\pi$.

PROPOSITION 13: *Consider the two incentive problems* $(A, U, \overline{U}, \pi, q)$, $(A, U, \overline{U}, \pi', q')$ *and assume that Assumptions* A1–A3 *hold for both. Suppose that* $\pi'(a) = R\pi(a)$ *for all* $a \in A$, *where* R *is an* $(n \times n)$ *stochastic matrix. Then* $C'(a) \geq C(a)$ *for all* $a \in A$. *Furthermore, if* V *is strictly concave and* $R \gg 0$,[16] *then* $C_{FB}(a^*) > \min_{a \in A} C_{FB}(a)$ *and* $C(a^*) < \infty \Rightarrow C'(a^*) > C(a^*)$.

PROOF: Let $(I_1', \ldots, I_n')$ be the cost minimizing way of implementing a in the primed problem. Suppose that in the unprimed problem, the principal offers the agent the following *random* incentive scheme: for each i, if q_i is the outcome, an n-sided die will be thrown where the probability of side j coming up is r_{ji}, the (j,i)th element of R $(j = 1, \ldots, n)$. If side j then comes up, you get I_j'. With this random incentive scheme, the probability of the agent getting I_j' if he chooses a particular action is the same as in the primed problem. Therefore the agent's optimal action will be a. Furthermore, the principal's expected costs are the same as in the primed problem. This shows that the principal can implement a at least as cheaply in the unprimed problem as in the primed problem by using a random incentive scheme. The final part of the proof is to note that the principal can reduce his expected cost further and continue to implement a by offering the agent the perfectly certain utility level $v_i = \sum_{j=1}^{n} r_{ji} V(I_j')$ if the outcome is q_i rather than the above lottery. That is, there is a deterministic incentive scheme which is better for the principal than the above random incentive scheme. *Q.E.D.*

REMARK 8: The last part of the proof of Proposition 13 shows that it is never desirable under our assumptions for the principal to offer the agent an incentive scheme which makes his payment conditional on a particular outcome a lottery rather than a perfectly certain income.[17] This result may also be found in Holmstrom [7].

Note that if $\pi' = R\pi$ and $q'R = q$, the random variable q' will have the same mean as q. In this case the following is true:

COROLLARY 1: *Make the hypotheses of Proposition* 13. *If, in addition,* q' *is such that* $q'R = q$, *we have* $L' \geq L$.

PROOF: Obvious since $B'(a) = q'\pi'(a) = q'R\pi(a) = q\pi(a) = B(a)$.

[16] We use this notation to mean that every element of R is strictly positive.

[17] This result depends strongly on our Assumption A1 that attitudes to income risk are independent of action. In the absence of this assumption, random incentive schemes may be desirable.

In the case $n = 2$, the transformation $\pi \to \pi' = R\pi$ is easy to interpret. Take any two actions $a_1, a_2 \in A$, and consider the likelihood ratio vector $(\pi_1(a_1)/\pi_1(a_2), \pi_2(a_1)/\pi_2(a_2))$. Assume without loss of generality that $\pi_1(a_1)/\pi_1(a_2) \leq \pi_2(a_1)/\pi_2(a_2)$. Then it is easy to show that

$$(5.1) \qquad \left[\frac{\pi_1'(a_1)}{\pi_1'(a_2)}, \frac{\pi_2'(a_1)}{\pi_2'(a_2)} \right] \subset \left[\frac{\pi_1(a_1)}{\pi_1(a_2)}, \frac{\pi_2(a_1)}{\pi_2(a_2)} \right],$$

where $[x, y]$ is the interval between x and y. In other words, the likelihood ratio vector becomes less variable in some sense when the stochastic transform R is applied. In fact the converse to this is also true: if (5.1) holds, then there exists a stochastic matrix R such that $\pi' = R\pi$ (see Blackwell and Girshick [3]). When $n > 2$, a simple characterization of this sort does not seem to exist, however.

One might ask whether a converse to Proposition 13 holds. That is, suppose $C'(a) \geq C(a)$ for all $a \in A$ and all concave utility functions V. Does it follow that $\pi'(a) = R\pi(a)$ for all $a \in A$, for some stochastic R? A converse along these lines can in fact be established when $n = 2$. Whether it holds for $n > 2$, we do not know.

Corollary 1 gives us a simple way of generating worse and worse incentive problems: repeatedly apply stochastic transforms to π. Suppose that we do this using always the same stochastic transform R, when $R \gg 0$ and is invertible. That is, we consider a sequence of incentive problems $1, 2, \ldots,$ where in the mth problem $\pi_m(a) = R^{m-1}\pi(a)$ for all $a \in A$, and the gross profit vector q_m satisfies $q_m R^{m-1} = q$ (this has a solution since R is invertible). We know from Corollary 1 that L_m will be increasing in m. The next proposition says that in the limit the loss from not being able to observe the agent reaches its maximal level.

DEFINITION: Let $L^* = \max_{a \in A}(B(a) - C_{FB}(a)) - \max\{B(a') - C_{FB}(a') \mid a'$ minimizes $C_{FB}(a)$ on $A\}$.

Since $C(a') = C_{FB}(a')$ if a' minimizes $C_{FB}(a)$, L^* is an upper limit on the loss to the principal from being unable to observe the agent. The next proposition shows that as the information q reveals about a gets smaller and smaller, the principal loses control over the agent, i.e., the agent chooses the least-cost action.

PROPOSITION 14: *Consider the sequence of incentive problems* $(A, U, \overline{U}, \pi_m, q_m)$, $m = 1, 2, \ldots,$ *where* $\pi_m(a) = R^{m-1}\pi_1(a)$ *for all* $a \in A$, $q_m R^{m-1} = q_1$ *for some invertible stochastic matrix* $R \gg 0$. *Assume* A1, A2, *and* $\pi_{1i}(a) > 0$ *for all* $i = 1, \ldots, n$, *and* $a \in A$. *Then if* V *is not a linear function,* $\lim_{m \to \infty} L_m = L^*$.

PROOF: It suffices to show that $\lim_{m \to \infty} C(a^*) = \infty$ for all a^* with $C_{FB}(a^*) > \min_{a \in A} C_{FB}(a)$. Suppose not for some such a^*. Let $(I_{m1}, \ldots, I_{mn})$ be the cost minimizing way of implementing a^* in problem m. Then $\sum_i \pi_{mi}(a^*) I_{mi}$ and $\sum_i \pi_{mi}(a^*) V(I_{mi})$ are both bounded in m. It follows from Bertsekas [2] that the (I_{mi}) are bounded. Hence without loss of generality we may assume $I_{mi} \to I_i$ for

each i. It is easy to show that, since R is a strictly positive stochastic matrix, $\lim_{m\to\infty} R^{m-1} = R^*$ where R^* has the property that all of its columns are the same. Therefore $\lim_{m\to\infty}\pi_m(a) = R^*\pi_1(a) = \bar{\pi}$ is independent of a. But this means $\lim_{m\to\infty}\sum_i \pi_{mi}(a^*)V(I_{mi}) = \sum_i \bar{\pi}_i V(I_i) = \lim_{m\to\infty}\sum_i \pi_{mi}(a)V(I_{mi})$ for all $a \in A$. Hence the agent will prefer actions a with $C_{FB}(a) < C_{FB}(a^*)$ to a^*. This contradicts the assumption that the incentive scheme implements a^*. *Q.E.D.*

We turn now to a consideration of another factor which influences L: the agent's degree of risk aversion. Since no incentive problem arises when the agent is risk netural, but an incentive problem does arise when the agent is risk averse, one is led to ask whether L increases as the agent becomes more risk averse. One difficulty in answering this question in general is the following. The way one makes the agent more risk averse is to replace his utility function $U(I,a)$ by $H(U(I,a))$ where H is a real-valued, increasing, concave function. However, if U satisfies Assumption A1, then $H(U)$ will generally not. To get around this difficulty, we will confine our attention to the case where A is a subset of the real line, $V(I) = -e^{-kI}$, $G(a) = 0$, and $K(a) = e^{ka}$, i.e., the agent's utility function is $U(a,I) = -e^{-k(I-a)}$, where $k > 0$. Assume also that $\bar{U} = -e^{-k\alpha}$, i.e., the agent's outside opportunity is represented by the perfectly certain income α. An increase in risk aversion can then be represented simply by an increase in k.

Note that if the agent's utility function is $-e^{-k(I-a)}$ and $\bar{U} = -e^{-k\alpha}$, then $C_{FB}(a) = a + \alpha$, which is independent of k. Hence first best profits are independent of k.

PROPOSITION 15: *Consider the incentive problem* $(A, U, \bar{U}, \pi, q)$ *where* A *is a subset of the real line,* $U(a,I) = -e^{-k(I-a)}$, $\bar{U} = -e^{-k\alpha}$, *and* $k > 0$. *Assume* A3. *Write the loss from being unable to observe the agent as* $L(k)$. *Then* $\lim_{k\to 0} L(k) = 0$, $\lim_{k\to\infty} L(k) = L^*$.

PROOF: To show that $\lim_{k\to\infty} L(k) = L^*$, it suffices to show that $\lim_{k\to\infty} C(a^*, k) = \infty$ for all a^* with $C_{FB}(a^*) > \min_{a\in A} C_{FB}(a)$. Suppose not for some such a^*, and let $C_{FB}(a) < C_{FB}(a^*)$. Then if $(I_1, \ldots, I_n)$ implements a^*, we must have

$$-\left(\sum_i \pi_i(a^*)e^{-kI_i}\right)e^{ka^*} \geq -\left(\sum_i \pi_i(a)e^{-kI_i}\right)e^{ka}$$

($I_1, \ldots, I_n$ of course depend on k). Therefore,

$$(5.2) \qquad e^{k(a^*-a)} \leq \sum_i \pi_i(a)e^{-kI_i} \Big/ \sum_i \pi_i(a^*)e^{-kI_i}.$$

Now let $k \to \infty$. The LHS of (5.2) $\to \infty$. Therefore so must the RHS. We may assume w.l.o.g., however, that $I_1 = \min_i I_i$. Then

$$\frac{\sum_i \pi_i(a)e^{-kI_i}}{\sum_i \pi_i(a^*)e^{-kI_i}} = \frac{\sum_i \pi_i(a)e^{k(I_1-I_i)}}{\sum_i \pi_i(a^*)e^{k(I_1-I_i)}},$$

which is bounded since the denominator $\geq \pi_1(a^*)$. Contradiction.

We show now that $\lim_{k\to 0} L(k) = 0$. Let $I_i = q_i - F$. Then the agent maximizes

$$(5.3) \qquad E(-e^{-k(I-a)}) = -E\left(1 - k(I-a) + \frac{k^2}{2}(I-a)^2 + \dots\right)$$

$$= -1 + k\left(\sum \pi_i(a) q_i - F - a\right) - \frac{k^2}{2} E(I-a)^2 + \cdots .$$

It follows that the agent maximizes

$$\left(\sum \pi_i(a) q_i - F - a\right) - \frac{k}{2} E(I-a)^2 + \cdots ,$$

which means that in the limit $k \to 0$ the agent maximizes $B(a) - C_{FB}(a)$, i.e. chooses a first-best action. Furthermore, setting (5.3) equal to $-e^{-k\alpha} = -1 + k\alpha + \cdots$, we see that in the limit $k \to 0$,

$$\max_{a \in A}\left(\sum_i \pi_i(a) q_i - a\right) - F = \alpha,$$

so that the principal's expected profit equals $F = \max_{a\in A}(\sum_i \pi_i q_i - a) - \alpha$ $= \max_{a\in A}(B(a) - C_{FB}(a)) =$ first-best profit. *Q.E.D.*

Proposition 15 tells us about the behavior of $L(k)$ for extreme values of k. It would be interesting to know whether $L(k)$ is increasing in k. We do not know the answer to that question except for the case $n = 2$, A finite.

PROPOSITION 16: *Make the same hypotheses as in Proposition* 14. *Assume in addition that* $n = 2$ *and* A *is finite. Then* $L(k)$ *is increasing in* k.

PROOF: See Appendix.

REMARK 9: Propositions 15 and 16 tell us how the principal's welfare varies with k. It is also interesting to ask how the shape of the optimal incentive scheme depends on k. Unfortunately, even in the case $n = 2$, very little can be said. In this case, the incentive scheme is characterized by the agent's share s. It is not difficult to construct examples showing that an increase in the agent's risk aversion may increase the optimal value of s, or may decrease it.

We conclude this section by considering how L depends on the agent's incremental costs. Consider the case of additive separability, i.e., $K(a) \equiv$ constant. Suppose that we write the agent's utility function as $U_\lambda(a, I) = G_\lambda(a) + V(I)$, where $G_\lambda(a) = \alpha + \lambda F(a)$, $\lambda > 0$. (Without loss of generality, we take $K = 1$.) Then, when λ is small, one feels that L will be small since the agent does not require much of a reward to work hard. The fact that $\lim_{\lambda\to 0} L(\lambda) = 0$ has in fact been established by Shavell [**20**]. We prove a somewhat stronger result.

PROPOSITION 17: *Consider the incentive problem* $(A, U_\lambda, \overline{U}, \pi, q)$, *where* $U_\lambda(a, I) = \alpha + \lambda F(a) + V(I)$ *for all* $a \in A$, $\lambda > 0$. *Assume that* A1–A3 *hold for this*

problem. Assume also that (1) *A is an interval of the real line*; (2) *$B(a)$ and $F(a)$ are twice differentiable in the interior of A*; (3) *V is twice differentiable on $\mathcal{S}$ and $V' > 0$*; (4) *There is a unique maximizer a^* of $B(a)$ lying in the interior of A and $B''(a^*) < 0$. Then* $\lim_{\lambda \to 0}(L(\lambda)/\lambda) = 0$.

PROOF: Consider the incentive problem with $\lambda = 1$. Then there are a's arbitrarily close to a^* for which $C(a)$ is finite. For let the principal set $v_i = rq_i - k$ where k is chosen so that $v_i \in \mathcal{R}$ for all i. Then the agent will maximize $\sum \pi_i(a) U_\lambda(a, I_i)$, i.e. $\sum \pi_i(a) q_i + F(a)/r$. By letting $r \to \infty$, we can get the agent to choose an action arbitrarily close to a^*. For such an action, $C(a)$ will be finite.

Consider now an a arbitrarily close to a^*. Let $(v_1, \ldots, v_n)$ be the cost minimizing way of implementing a when $\lambda = 1$. Then it is clear from (2.2) that $(\lambda v_1 + \beta, \ldots, \lambda v_n + \beta)$ will implement a for $\lambda \neq 1$, where

$$\lambda\Big(\sum \pi_i(a) v_i + F(a)\Big) + \alpha + \beta = \bar{U}.$$

It follows that

$$L(\lambda) \leq \sum \pi_i(\hat{a}) q_i - h\Big(\bar{U} - \alpha - \lambda F(\hat{a})\Big) - \Big(\sum \pi_i(a) q_i - \sum \pi_i(a) h(\lambda v_i + \beta)\Big),$$

where $\hat{a}$ maximizes $\sum \pi_i(a) q_i - h(\bar{U} - \alpha - \lambda F(a))$, i.e. $\hat{a}$ is the first-best action in problem λ.

Therefore,

$$\frac{L(\lambda)}{\lambda} \leq \left[\frac{1}{\lambda}\left\{\sum \pi_i(a^*) q_i - h\Big(\bar{U} - \alpha - \lambda F(a^*)\Big) - \Big(\sum \pi_i(a) q_i - \sum \pi_i(a) h(\lambda v_i + \beta)\Big)\right\}\right] + \left[\frac{1}{\lambda}\left\{\sum \pi_i(\hat{a}) q_i - h\Big(\bar{U} - \alpha - \lambda F(\hat{a})\Big) - \sum \pi_i(a^*) q_i + h\Big(\bar{U} - \alpha - \lambda F(a^*)\Big)\right\}\right].$$

Now $\hat{a} \to a^*$ as $\lambda \to 0$. Furthermore, by differentiating the first-order conditions $(d/da)(\sum \pi_i(\hat{a}) q_i - h(\bar{U} - \alpha - \lambda F(\hat{a})) = 0$, one can show that $d\hat{a}/d\lambda$ exists at $\lambda = 0$. It follows from the mean-value theorem and the fact that $B'(a^*) = 0$ that the second square bracket $\to 0$ as $\lambda \to 0$. To see that the first square bracket $\to 0$, note that, since a is arbitrary, we can make a converge to a^* as fast as we like. Therefore we need only show that

$$\lim_{\lambda \to 0} \frac{1}{\lambda}\Big(\sum \pi_i(a) h(\lambda v_i + \beta) - h\Big(\bar{U} - \alpha - \lambda F(a^*)\Big)\Big) = 0. \tag{5.4}$$

But

$$\sum \pi_i(a)\Big[h(\lambda v_i + \beta) - h\big(\bar{U} - \alpha - \lambda F(a^*)\big)\Big]$$

$$= \sum \pi_i(a)\Big[h\big(\lambda v_i + \bar{U} - \alpha - \lambda \sum \pi_j(a) v_j - \lambda F(a)\big)$$

$$- h\big(\bar{U} - \alpha - \lambda F(a^*)\big)\Big]$$

$$= \sum \pi_i(a)\Big[h(\bar{U} - \alpha) + h'(\bar{U} - \alpha)\big(\lambda v_i - \lambda \sum \pi_j(a) v_j - \lambda F(a)\big)$$

$$+ \cdots - h(\bar{U} - \alpha) + h'(\bar{U} - \alpha)(\lambda F(a^*)) + \cdots\Big]$$

$$= h'(\bar{U} + \alpha)(-\lambda F(a) + \lambda F(a^*)) + \cdots$$

from which (5.4) follows. *Q.E.D.*

The proof of Proposition 17 is based on an envelope argument. It appears that a similar result can be established for the more general case where U is not additively separable, but Assumption A1 holds. Since the proof is more complicated, however, we will not pursue this result here. The assumption that a^* lies in the interior of A may seem quite strong. Note, however, that if a^* is a boundary point and $B'(a^*) \neq 0$, then the second-best optimal action equals a^* for small enough λ. It is straightforward to apply the proof of Proposition 17 to show that $\lim_{\lambda \to 0}(L(\lambda)/\lambda) = 0$ in this case too.

Since the marginal product of labor of the agent—that is, the increase in expected profit resulting from an extra pound of expenditure by the agent—is proportional to $1/\lambda$, Proposition 17 can be interpreted as saying that the welfare loss L is of a smaller order of magnitude than the reciprocal of the agent's marginal product of labor.

6. EXTENSIONS

We have assumed throughout the paper that the principal is risk-neutral and that the agent's attitudes to risk over income lotteries are independent of action —Assumption A1. We now briefly consider what happens if we relax these assumptions.

As we have noted in Section 2, Remark 3, our method of analysis generalizes without any difficulty to the case where the principal is risk-averse. Specific results change, however, The main difference is that now, even in the first-best situation, the principal will not bear all the risk. One implication of this is that even if there is no disutility of action for the agent, i.e. a does not enter the agent's utility function, the first-best will not generally be reached. The reason is that there may be a conflict between the principal and agent over what income lottery should be selected (for a study of this conflict, see Ross [**17**] and Wilson [**23**]).

As a result of this, Proposition 3, part (5), is no longer true when the principal is risk-averse. Nor is Proposition 17 since $L(0) \neq 0$. Propositions 1 and 2 and Proposition 3, parts 1–4, continue to hold, however. So do Propositions 4 and 5 on the characterization of an optimal incentive scheme. Propositions 7, 8 generalize, as do Propositions 10, 11, and 12 (note that the function C_{FB} is still well defined although it no longer refers to first-best cost). Proposition 3(6) does not hold and neither does Proposition 6 nor Proposition 9 (at least in its present form). Finally Corollary 1 of Proposition 13 and Propositions 14–16 do not generalize in an obvious way, since changing the risk aversion of the agent or the probability distribution of outcomes affects the first-best as well as the second-best.

The computational procedure presented in Section 4 for the two outcome case can be extended to the case where the principal is risk-averse. In the finite action case, it is still true that the agent will be indifferent between two actions at the optimum, except in the case where the first-best can be achieved. Thus it is necessary to check whether the first-best can be achieved. Otherwise the procedure is unaltered.

We turn now to the consequences of relaxing Assumption A1. These are much more serious since most of our analysis has depended crucially on being able to choose the control variables $V(I_1), \ldots, V(I_n)$ independently of a. Some results do generalize, however. In particular one can show that Propositions 1, 3, 10, and 12 generalize. It seems unlikely that the characterization of an optimal incentive scheme in Proposition 4 and Proposition 5, part 1, holds, but we do not have a counterexample. Surprisingly, perhaps, Proposition 5, part 2 does hold. Proposition 6 does not hold and it seems unlikely that Propositions 7–9 do.

In the two outcome case, one can still show that it is optimal for the agent's share s to satisfy $0 \leq s < 1$. As a consequence Propositions 10 and 12 generalize. Proposition 11 does not generalize, however, and nor does our computational procedure for the two outcome case. Propositions 13 and 14 and Corollary 1 of Proposition 14 do not hold as they stand, although they do if one enlarges the set of feasible incentive schemes to include random schemes. (As we have noted in footnote 17, once Assumption A1 is dropped, random incentive schemes may be superior to deterministic schemes.) Finally, it seems likely that Proposition 17 could be generalized to the nonseparable case.

7. SUMMARY

The purpose of this paper has been to develop a method for analyzing the principal-agent problem in the case where the agent's attitudes to income risk are independent of action. Our method consists of breaking up the principal's problem into a computation of the costs and benefits accruing to the principal when the agent takes a particular action. We have used this method to establish a number of results about the structure of the optimal incentive scheme and about the determinants of the welfare loss resulting from the principal's inability to observe the agent's action. We have shown that it is never optimal for the

incentive scheme to be such that the principal's and agent's payoff are negatively related over the whole outcome range, although such a relationship may be optimal over part of the range. We have found sufficient conditions for the incentive scheme to be monotonic, progressive, and regressive. We have shown that a decrease in the quality of the principal's information in the sense of Blackwell increases welfare loss. When there are only two outcomes, welfare loss also increases when the agent becomes more risk averse. Finally, we have discussed how our techniques can be used to compute optimal incentive schemes in particular cases.

While we have talked throughout about "the" principal-agent problem, we have in fact been considering the simplest of a number of such problems. More complicated principal-agent problems arise when not only is the principal unable to monitor the agent, but also the agent possesses information about his environment, i.e. about A, π, or $U(a, I)$, which the principal does not. Such problems possess a number of features of the preference revelation problems studied in the recent incentive compatibility literature; see, for example, the *Review of Economic Studies* Symposium [**16**]. A start has been made in the analysis of such problems by Harris and Raviv [**6**], Holmstrom [**7**], and Mirrlees [**12**]. It will be interesting to see whether the techniques presented here will also be useful in the solution of these more complicated principal-agent problems.

University of Chicago
and
London School of Economics

Manuscript received September, 1980; revisions received September, 1981.

APPENDIX

PROOF OF PROPOSITION 16: It suffices to show that $C(a, k)$ is increasing locally in k for each $a \in A$ whenever $C(a, k)$ is finite. Let $\tilde{k} = \lambda k$ $\lambda \geq 1$. Assume that (I_1, I_2) is the cost minimizing way of implementing a, given $\tilde{k}$. Then, by the results of Section 4, e.g. equation (4.4),

$$\pi_1 w_1 + \pi_2 w_2 = \frac{1}{e^{\tilde{k}(a+\alpha)}},$$

$$\pi_1' w_1 + \pi_2' w_2 = \frac{1}{e^{\tilde{k}(a'+\alpha)}}, \tag{A1}$$

where $w_1 = e^{-\tilde{k}I_1}$, $w_2 = e^{-\tilde{k}I_2}$, $\pi_1 = \pi_1(a)$, $\pi_2 = \pi_2(a)$, $\pi_1' = \pi_1(a')$, $\pi_2' = \pi_2(a')$, $a' \in A$, $a' < a$. Furthermore we can pick a' so that a' is independent of $\tilde{k}$ for λ close to 1.

Equations (A1) determine w_1 and w_2 for each value of $\tilde{k}$. The cost of implementing a, $C(a, \tilde{k})$, is then given by

$$C(a, \tilde{k}) = \pi_1 I_1 + \pi_2 I_2 = -\frac{1}{\tilde{k}}(\pi_1 \log w_1 + \pi_2 \log w_2). \tag{A2}$$

Differentiating (A2) with respect to λ we get

$$\left.\frac{\partial C(a, \lambda k)}{\partial \lambda}\right|_{\lambda=1} = \frac{1}{k}\left(\pi_1 \log w_1 + \pi_2 \log w_2 - \frac{\pi_1}{w_1}\frac{dw_1}{d\lambda} - \frac{\pi_2}{w_2}\frac{dw_2}{d\lambda}\right). \tag{A3}$$

Set $x = e^{-k(a+\alpha)}$, $y = e^{-k(a'+\alpha)}$ in (A1). Then $e^{-\tilde{k}(a+\alpha)} = x^{\lambda}$, $e^{-\tilde{k}(a'+\alpha)} = y^{\lambda}$. Hence

$$\text{(A4)} \qquad \begin{aligned} \pi_1 \frac{dw_1}{d\lambda} + \pi_2 \frac{dw_2}{d\lambda} &= x \log x, \\ \pi_1' \frac{dw_1}{d\lambda} + \pi_2' \frac{dw_2}{d\lambda} &= y \log y, \end{aligned}$$

where derivatives are evaluated at $\lambda = 1$. Solving (A1), (A4) yields

$$w_1 = \frac{\pi_2' x - \pi_2 y}{\pi_1 \pi_2' - \pi_1' \pi_2} = \frac{\pi_2' x - \pi_2 y}{\pi_2' - \pi_2},$$

$$\frac{dw_1}{d\lambda} = \frac{\pi_2' x \log x - \pi_2 y \log y}{\pi_2' - \pi_2}.$$

It follows that $\log w_1 \geq (1/w_1)(dw_1/d\lambda)$. For

$$\text{(A5)} \qquad w_1 \log w_1 - \frac{dw_1}{d\lambda} = \frac{\pi_2' x - \pi_2 y}{\pi_2' - \pi_2} \log \frac{\pi_2' x - \pi_2 y}{\pi_2' - \pi_2} - \left(\frac{\pi_2' x \log x - \pi_2 y \log y}{\pi_2' - \pi_2} \right)$$

$$= \frac{1}{\pi_2' - \pi_2} \left[(\alpha x - \beta y) \log \frac{\alpha x - \beta y}{\alpha - \beta} - \alpha x \log x - \beta y \log y \right],$$

where $\alpha = \pi_2'$, $\beta = \pi_2$. However, the RHS of (A5) ≥ 0 by Lemma 3 below. The same argument shows that $\log w_2 \geq (1/w_2)(dw_2/d\lambda)$. It follows from (A3) that $(\partial C/\partial\lambda) \geq 0$, i.e., C is increasing locally in k.

LEMMA 3: *Assume* $\alpha, \beta, x, y > 0$. *Then if* $\alpha > \beta$ *and* $\alpha x > \beta y$, $\alpha x \log x - \beta y \log < (\alpha x - \beta y) \log((\alpha x - \beta y)/(\alpha - \beta))$. *On the other hand, if* $\alpha < \beta$ *and* $\alpha x < \beta y$, $\alpha x \log x - \beta y \log y > (\alpha x - \beta y) \log((\alpha x - \beta y)/(\alpha - \beta))$.

PROOF: Since $z \log z$ is a convex function,

$$\frac{\beta}{\alpha}(y \log y) + \left(\frac{\alpha - \beta}{\alpha} \right) \left(\frac{\alpha x - \beta y}{\alpha - \beta} \log \frac{\alpha x - \beta y}{\alpha - \beta} \right) \geq x \log x.$$

This proves the first part. The second part follows similarly. *Q.E.D.*

REFERENCES

[1] ARROW, K. J.: "Insurance, Risk and Resource Allocation," *Essays in the Theory of Risk Bearing*. Chicago: Markham, 1971.

[2] BERTSEKAS, D.: "Necessary and Sufficient Conditions for Existence of an Optimal Portfolio," *Journal of Economic Theory*, 8(1974), 235–247.

[3] BLACKWELL, D., AND M. A. GIRSHICK: *Theory of Games and Statistical Decisions*. New York: John Wiley and Sons, Inc., 1954.

[4] BORCH, K.: *The Economics of Uncertainty*. Princeton: Princeton University Press, 1968.

[5] HARDY, G. H., J. E. LITTLEWOOD, AND G. POLYA: *Inequalities*. Cambridge: Cambridge University Press, 1952.

[6] HARRIS, M., AND A. RAVIV: "Optimal Incentive Contracts with Imperfect Information," *Journal of Economic Theory*, 20(1979), 231–259.

[7] HOLMSTROM, B.: "Moral Hazard and Observability," *Bell Journal of Economics*, 10(1979), 74–91.

[8] KEENEY, R.: "Risk Independence and Multiattributed Utility Functions," *Econometrica*, 41(1973), 27–34.

[9] MILGROM, P. R.: "Good News and Bad News: Representation Theorems and Applications," Discussion Paper No. 407, Northwestern University, Illinois, Mimeo, 1979.

[10] MIRRLEES, J. A.: "The Theory of Moral Hazard and Unobservable Behavior—Part I," Nuffield College, Oxford, Mimeo, 1975.
[11] ———: "The Optimal Structure of Incentives and Authority Within an Organization," *Bell Journal of Economics*, 7(1976), 105–131.
[12] ———: "The Implications of Moral Hazard for Optimal Insurance," Seminar given at Conference held in honour of Karl Borch, Bergen, Norway, Mimeo, 1979.
[13] PAULY, M.: "The Economics of Moral Hazard: Comment," *American Economic Review*, 58(1968), 531–536.
[14] POLLAK, R.: "The Risk Independence Axiom," *Econometrica*, 41(1973), 35–39.
[15] RADNER, R.: "Monitoring Cooperative Agreements in a Repeated Principal-Agent Relationship," Mimeo, Bell Laboratories, 1980.
[16] *Review of Economic Studies* Symposium on Incentive Compatibility, April, 1979.
[17] ROSS, S.: "The Economic Theory of Agency: The Principal's Problem," *American Economic Review*, 63(1973), 134–139.
[18] RUBINSTEIN, A. AND M. YAARI: Seminar given at Conference held in honour of Karl Borch, Bergen, Norway, 1979.
[19] SHAVELL, S.: "On Moral Hazard and Insurance," *Quarterly Journal of Economics*, 93(1979), 541–562.
[20] ———: "Risk Sharing and Incentives in the Principal and Agent Relationship," *Bell Journal of Economics*, 10(1979), 55–73.
[21] SPENCE, M., AND R. ZECKHAUSER: "Insurance, Information, and Individual Action," *American Economic Review*, 61(1971), 380–387.
[22] STIGLITZ, J. E.: "Incentives and Risk Sharing in Sharecropping," *Review of Economic Studies*, 61(1974), 219–256.
[23] WILSON, R.: "The Theory of Syndicates," *Econometrica*, 36(1968), 119–132.
[24] ZECKHAUSER, R.: "Medical Insurance: A Case Study of the Trade-Off Between Risk Spreading and Appropriate Incentives," *Journal of Economic Theory*, 2(1970), 10–26.