CHAPTER III

COORDINATION OF INFORMATION

3.1 Introduction

In Chapter II we studied extensively a special type of decentralization called delegation. The characteristic feature of delegation is that no coordination of information takes place. As a result, agents need not communicate their information to the center, which normally means substantial savings in information costs. Another advantage is that each agent's reward is independent of the other agents' actions (unless there are externalities); generally viewed as a desirable organizational feature. The drawback, of course, is that one foregoes the opportunities of improved decision making via coordination of information.

In this chapter we will discuss more general decentralization procedures, which coordinate the information of agents. Our interest lies in studying the possibilities of achieving efficient outcomes in each informational state.[1] The basic model was introduced in Chapter I. As we argued there, efficient outcomes can be achieved if and only if agents will tell the truth when an efficient decision function is employed. If an efficient decision function induces truth-telling, the corresponding decision mechanism is said to be incentive compatible (i.c.), if truth-telling will be a dominant strategy for each agent,

then the mechanism is called strongly incentive compatible (s.i.c.); if truth-telling will be an ordinary Nash-equilibrium we say that the mechanism is weakly incentive compatible (w.i.c.). Of course, s.i.c. implies w.i.c.

We will mainly restrict ourselves to a discussion of Groves' scheme. In a path-breaking paper [1973], Groves showed that if agents have preference functions which are additively separable and linear in money, then there exists a set of monetary compensation rules, based on messages alone, which will induce agents to tell the truth. In his original formulation, Groves' scheme was only w.i.c., because of the possibility of partial communication of information. Later on, Loeb [1975] and Groves and Loeb [1975] showed that the scheme became s.i.c. with full communication. Since then, much research has been devoted to analyzing properties of Groves' scheme; particularly in the framework of full communication. Among other things, it has been proved (Green and Laffont [1977]) that Groves' scheme is essentially the unique s.i.c. mechanism under the assumption of a universal domain of preferences.

Our main results are further uniqueness characterizations. First, we show that the universal domain assumption can be dispensed with if a weak differentiability condition is satisfied. Secondly, we give an analogous characterization for w.i.c. mechanisms, which states that every w.i.c. mechanism equals a Groves' mechanism in expectation. So even when only limited communication is allowed, Groves' scheme is essentially unique.

Groves' scheme has some undesirable features, which also have received attention. A serious problem is the fact that the monetary compensations may not net to zero. If they generate a deficit, the scheme is effectively infeasible; if they generate a surplus, it is wasteful, and full efficiency is not achieved. We will derive a necessary and sufficient condition under which the compensation rules can be chosen so that the budget balances. We will also discuss some other remedies to the problem; in particular, the work of d'Aspremont and Gerard-Varet [1975], who show that the budget can always be balanced when the agents' information is independent and only w.i.c. is required.

The technique we use to prove our uniqueness theorems is valuable in revealing the simple rationale behind Groves' scheme. This insight is important when trying to construct i.c. mechanisms in more complex environments. We will discuss such extensions in the last section. There we show that in a syndicate with members that have exponential utility functions, revelation of risk-tolerances can be induced so that efficient risk-sharing and decision-making can be achieved. A negative result, which has been found by Hurwicz [1972], states that efficient outcomes cannot generally be attained in exchange economies. We will show this result in a two-person case, using a new argument which does not rely on Hurwicz's assumption of individual rationality.

The outline of the chapter is as follows: Section 3.2 discusses strong incentive compatibility; Section 3.3, weak incentive compatibility; and Section 3.4, extensions. Section 3.5 contains concluding remarks.

3.2 Strong Incentive Compatibility

3.2.1 A Simple Model

We start with a simple model. The purpose is to show how Groves' scheme can be easily derived from the two conditions that the decision function is efficient and that truth-telling is a dominant strategy for the agents. Until Section 3.4, we will work under the assumption that agents have preference functions which are separable and linear in money; i.e., they can be written:

$$f_i(d,y_i) + t_i, \qquad i = 1, \ldots, n. \tag{3.1}$$

d is a decision to be made, the signal y_i is a parameter of agent i's utility function, and t_i is a monetary compensation by which the principal can change agent i's incentives to communicate information properly.

The principal's objective is to attain an efficient outcome function d*(y). The form of the utility functions in (3.1) imply that d*(y) satisfies:

$$d^*(y) = \underset{d}{\text{argumax}} \sum_{i=1}^{n} f_i(d,y_i), \qquad \forall y. \tag{3.2}$$

To achieve efficiency, the principal asks the agents for their private information y, and selects d*(m) as his response to their messages $m = (m_1, \ldots, m_n)$. The agents will tell the truth as a dominant strategy -- and consequently the outcome function d*(y) can be

attained -- if and only if there exist compensation rules $t_i(m)$ such that:[2]

$$(3.3) \qquad y_i = \underset{m_i}{\text{argmax}} \; [f_i(d^*(m_i, y^i), y_i) + t_i(m_i, y^i)], \quad \forall y, \forall i.$$

Notice that we have written y^i rather than m^i in (3.3). The reason is that it does not matter to agent i whether the others tell the truth or not, because of the form of his preference function. In particular, it is a consequence of the fact that his preference function does not depend directly on y^i.[3] If f_i depended on y^i, then we could not hope for a dominant strategy. We will return to this point later.

Assume now that $y_i \in R^1$, $\forall i$, and that $f_i(d^*(m_i, y^i), y_i)$ and $t_i(m_i, y^i)$ are differentiable w.r.t. m_i, $\forall i$. From (3.2) and (3.3) follows:

$$(3.4) \qquad \frac{\partial}{\partial m_i} \sum_{j=1}^{n} f_j(d^*(y), y_j) = 0,$$

$$(3.5) \qquad \frac{\partial}{\partial m_i} [f_i(d^*(y), y_i) + t_i(y)] = 0,$$

for every y. Substitution of (3.5) into (3.4) yields:

$$\frac{\partial}{\partial m_i} t_i(y) = \frac{\partial}{\partial m_i} \sum_{j \neq i} f_j(d^*(y), y_j),$$

which integrates to the family of solutions:

$$(3.6) \qquad t_i(y) = \sum_{j \neq i} f_j(d^*(y), y_j) + h_i(y^i).$$

The solution is unique up to the arbitrary function $h_i(y^i)$, which is independent of y_i.

The compensation rules defined by (3.6) for every i is called Groves scheme, and will be denoted by $g_i(m_i; h_i)$, $i = 1, \ldots, n$. We have shown, in this simple context, that a compensation rule which induces truth-telling must be a Groves' scheme. On the other hand, if Groves' scheme is used, then the agent and the principal will have identical objectives (when viewed as functions of m_i alone). Both want to maximize:

$$(3.7) \qquad \sum_{j=1}^{n} f_j(d^*(m_i, y^i), y_j), \text{ for every } y.$$

Consequently, truth-telling will be a dominant strategy. By telling the truth, the agent lets the principal solve his problem; or put in another way, the agent would select the same decision as the principal would, if the agent knew y^i.

Though the derivation above was carried out in a simple model, it reveals the essentials of the problem structure and the reason why Groves' scheme is the only compensation rule that can achieve an efficient decision (excluding monetary compensations) when preferences are independent and linearly separable in money. The point is that each agent must carry the total social cost of changing the decision,

in order for him to behave in a socially optimal way. The separability of his preference function allows us to impose the total social cost on him.

3.2.2 Optimality and Uniqueness of Groves' Scheme

The foregoing derivation of Groves' scheme can easily be extended. We assumed one-dimensional messages for notational simplicity, and the differentiability assumptions were unnecessarily strong. We proceed now to a more general statement of the problem.

Let $d \in D$ be the joint decision, and let each agent in addition make a local decision $a_i \in A_i$, which does not enter the other agents' preference functions. The preference functions are assumed additively separable and linear in money; i.e., of the form:

$$F_i(d,a_i,z) + t_i, \qquad i = 1, \ldots, n. \tag{3.8}$$

z is the state of nature. Agents observe signals $y_i \in \mathbb{R}^{k_i}$, $i = 1, \ldots, n$, which may be characteristics about their preferences as well as information about some future events that affect their utilities. These signals are assumed conditionally independent in the following sense:

$$E\{F_i(d,a_i,z)|y\} = E\{F_i(d,a_i,z)|y_i\}, \quad \forall y, \forall i. \tag{3.9}$$

In other words, knowing the other agents' signals does not change agent i's expected utility (see p. 153 for a discussion of this condition).

Define the derived preference functions:

$$(3.10) \qquad f_i(d,y_i) = \max_{a_i \in A_i} E\{F_i(d,a_i,z)|y_i\}, \quad \forall i.$$

Since the agents make their local decisions after the joint decision d is made, the derived preference functions are the only ones of relevance for the principal in choosing d. We assume that the principal knows the functional form of the f_i's, but not the parameter values.

For each y, we assume there exists a Pareto optimal decision denoted d*(y). Because of the form of the preference functions (see (3.8)) it has to satisfy:

$$(3.11) \qquad d^*(y) = \operatorname*{argmax}_{d \in D} \sum_{i=1}^{n} f_i(d,y_i) - c(d),$$

where c(d) is an external social cost of the decision d. Agents report their full information to the principal, who uses d*(•) as his decision function. We are looking for monetary transfers $t_i(m)$, which make truth-telling a dominant strategy for each agent; that is, which satisfy:

$$(3.12) \qquad y_i = \operatorname*{argmax}_{m_i \in \mathbb{R}^{k_i}} [f_i(d^*(m_i,y^i),y_i) + t_i(m_i,y^i)], \quad \forall y, \forall i.$$

Define the following social objective function:

$$S(m;y) = \sum_{i=1}^{n} f_i(d^*(m),y_i) + c(d^*(m)).$$

We will assume that S is differentiable w.r.t. m at y, for each y. No differentiability assumptions are made on transfers. By definition of $d^*(y)$ in (3.11), it follows that $S(m;y)$ is maximized at $m = y$, implying:

$$\nabla S(y;y) = 0, \tag{3.13}$$

since the message space is open.

We will show that under the assumptions made above, Groves' scheme is the unique scheme which achieves efficiency. We start with a lemma, which shows why no differentiability assumptions need to be made for the t_i's.[4]

Lemma 3.1: Let $f = \mathbb{R}^{2k} \to \mathbb{R}$, $t = \mathbb{R}^k \to \mathbb{R}$ and assume:

(i) $y = \underset{m \in \mathbb{R}^k}{\operatorname{argmax}} f(m,y), \quad \forall y \in \mathbb{R}^k,$

(ii) $y = \underset{m \in \mathbb{R}^k}{\operatorname{argmax}} [f(m,y) + t(m)], \quad \forall y \in \mathbb{R}^k,$

(iii) f is differentiable w.r.t. m, at y for every y.

Then t(m) = constant.

Proof: Suppose t is not constant. Then it follows that there exists an $\bar{m} \in \mathbb{R}^k$, a sequence $\{m^\ell\}$ converging to $\bar{m}$, and an $\varepsilon > 0$, such that

(3.14) $$t(m^{\ell}) - t(\bar{m}) > \varepsilon \cdot |m^{\ell} - \bar{m}|.$$

(If not t would be differentiable at each m with a gradient equal to 0.) From (i) and (iii) we know that,

$$0 < f(\bar{m},\bar{y}) - f(m^{\ell},\bar{y}) = o(m^{\ell} - \bar{m}),$$

where $\bar{y} = \bar{m}$. Consequently, for small $|m^{\ell} - \bar{m}|$, the difference in the t-function will dominate the difference in the f-function, and (3.14) will contradict (ii). Q.E.D.

Using the lemma we can prove the main uniqueness theorem:

Theorem 3.2: If the social objective function S(m;y) is differentiable w.r.t. m at y, for each y, then truth-telling will be a dominant strategy if and only if the compensation rules $t_i(\cdot)$, $i = 1, \ldots, n$ are Groves' schemes, i.e.:

(3.15) $$t_i(m) = g_i(m;h_i) \equiv \sum_{j \neq i} f_j(d^*(m),m^i) - c(d^*(m)) + h_i(m^i), \quad \forall i,$$

for some choice of h_i-functions.

Proof: The if-part has been proved by Groves and Loeb [1975] and is an immediate consequence of the fact that the principal's

objective function $S(m;y)$ coincides with the agent's, if Groves' scheme is used (cf. (3.7)).

To prove the only-if part, let

$$h_i(y) = t_i(y) - \sum_{j\neq i} f_j(d^*(y),y_j) + c(d^*(y)).$$

Agent i's objective function can then be written

$$\sum_{j=1}^{n} f_j(d^*(m_i,y^i),y_j) - c(d^*(m_i,y^i)) + h_i(m_i,y^i),$$

when the others report y^i. We need to show that $h_i(m_i,y^i)$ is independent of m_i. Keeping y^i fixed, this follows directly from Lemma 3.1, since assumption (i) is satisfied by definition of d^* (see (3.11)); (ii) is satisfied since we assumed truth-telling is a dominant strategy; and (iii) is our differentiability assumption. Q.E.D.

Remark: Differentiability is, of course, only needed for the uniqueness characterization.

Theorem 3.2 is essentially the uniqueness result of Green and Laffont [1977]. The difference is that they assume a universal domain for the f_i-functions, whereas we have parametrized these functions in $\mathbb{R}^{k_i}$, respectively. In order to prove uniqueness in such restricted domains, we had to assume differentiability of the social welfare function. This is crucial when the parameter is one-dimensional, but

as the dimensionality of the message space is increased, it is likely that the differentiability condition could be weakened. We have not studied how, but the conjecture is naked, since we know that no differentiability is needed when the domain is unrestricted.

We notice in this connection that the proof above could have been extended to wider parameter spaces than $\mathbb{R}^k$. For instance, y_i could be an element in some function space. The corresponding differentiability assumption would be that the social objective function $S(m;y)$ has a derivative in each direction (i.e., it is Gateaux-differentiable; see Luenberger [1968]).

We want to emphasize that our differentiability assumption is rather weak. By no means does it imply that the social welfare function $(\sum_{i=1}^{n} f_i(d,y_i) - c(d))$ has to be differentiable w.r.t. d. In fact, d may very well take on only a finite number of values, or belong to a compact set with boundaries binding for some values of y. This can be illustrated by studying the problem of accepting or rejecting a public project.

Example 3.1: A 0-1 Public Project Problem.[5]

The principal (e.g., the government) has to decide on undertaking a public project, say the construction of a bridge. Let $d = 0$ denote rejection of the project and $d = 1$, acceptance of it. In order to make a Pareto optimal decision, he asks the individuals in the society for their willingness to contribute to the cost of the project, which is c dollars. Let y_i be the true value to agent i of the project,

expressed in dollars, and m_i his reported willingness to pay for it. The Pareto optimal decision function is then:

$$(3.16) \qquad d^*(y) = 1, \quad \text{if } \sum_i y_i > c,$$

$$= 0, \quad \text{if } \sum_i y_i \leq c.$$

The social objective function is:

$$S(m;y) = \sum_i y_i - c, \quad \text{if } \sum_i m_i > c,$$

$$= 0, \quad \text{if } \sum_i m_i \leq c.$$

Here we have scaled the agents' utility functions so that they are zero when the project is not undertaken. The monetary compensations with Groves' scheme are:

$$t_i(m) = \sum_{j \neq i} m_j - c + h_i(m^i), \quad \text{if } \sum_{i=1}^{n} m_i \geq c,$$

$$= h_i(m^i), \quad \text{if } \sum_{i=1}^{n} m_i < c.$$

Despite its discontinuous nature as a function of d, it is readily checked that $S(m;y)$ is differentiable w.r.t. m at y, for each y. Indeed, for every y, it is constant in a neighborhood of $m = y$ (and when $\Sigma y_i = c$, $S = 0$). Hence, Theorem 3.2 is applicable, and we conclude that only Groves' scheme is strongly incentive compatible in this context.[6] □

From the example it is clear that whenever d takes on discrete values, and the preference functions are smooth functions of the parameters y, differentiability is guaranteed. Likewise, we find that our differentiability condition holds also for problems where the preference functions are smooth functions of the decision d, even if d is restricted to some compact set whose boundaries may be binding. From this we conclude that our differentiability assumption is quite generally satisfied.

Of course, uniqueness is lost if we restrict ourselves to discrete domains of the preference parameters.

It is easy to see why one needs an assumption about conditional independence, in order for Groves' scheme to work. Without the independence assumption, the social objective function would look as follows:

$$(3.17)\qquad S(m_i,y^i;y) = f_i(d^*(m_i,y^i),y) + \sum_{j\neq i} f_j(d^*(m_i,y^i),y) - c(d^*(m_i,y^i)),$$

written as a function of m_i. Here d^* is the Pareto optimal decision function, as before. If Groves' scheme is applied, the agent's objective function becomes:

$$(3.18)\qquad f_i(d^*(m_i,y^i),y) + \sum_{j\neq i} f_j(d^*(m_i,y^i),(m_i,y^i)) - c(d^*(m_i,y^i)),$$

provided the other agents report truthfully. Because y_i appears in the other agents' preference functions, (3.18) differs from (3.17) and the optimal message for the agent will no longer be y_i. For an illustration of this point, see Example 3.2.

In Section 3.4 we will give an example which shows that one can, at least sometimes, find other schemes which induce truth-telling when preferences are dependent. However, notice that generally such schemes cannot be expected to achieve strong incentive compatibility. If other agents lie, agent i would like to compensate for this in the choice of his message, because the principal will not act according to his interest.

3.2.3 Budget-Balancing

We have shown that Groves' scheme attains the efficient decision function $d^*(y)$ via dominant truth-telling strategies. But notice that the total social decision includes the transfer payments t_i, so it is false to say that Groves' scheme yields efficient outcomes, unless we can show that the pair $(d(y), t(y))$ can be chosen efficiently for all y. Because agents' preference functions are separable, we can determine independently the efficient decision function $d^*(y)$ and efficient transfer payments t_i. $d^*(y)$ is defined by (3.11), and transfer payments will be efficient (and feasible in the sense that they cover the cost c) if and only if,

$$(3.19) \qquad \sum_{i=1}^{n} t_i(y) = -c(d^*(y)), \quad \forall y.$$

This condition is known as budget-balancing.

Because we know from Theorem 3.2 that $d^*(y)$ can be attained only by using a Groves' scheme, it is of interest to ask when there exist Groves' transfer payments $g_i(y;h_i)$ such that (3.19) holds. Only then can we say that full efficiency is guaranteed using a Groves' mechanism.

To analyze this question, define the social net deficit function:

$$p(y) = -\sum_{i=1}^{n} f_i(d^*(y),y_i) + c(d^*(y)). \tag{3.20}$$

If agents report y as their preferences, then the principal has to pay out $(n-1) \cdot p(y)$, including the cost $c(d(y))$, but excluding $\sum_{i=1}^{n} h_i(y^i)$. We say that $p(y)$ is $(n-1)$-separable if we can write:

$$p(y) = \sum_{i=1}^{n} p_i(y^i),$$

where p_i is independent of y_i. We have:

Theorem 3.3: There exists a set of budget-balancing Groves' transfer payments if and only if $p(y)$ in (3.20) is $(n-1)$-separable.

Proof: Suppose $p(y)$ is $(n-1)$-separable. Choose

$$h_i(y^i) = -(n-1) \cdot p_i(y^i), \quad i = 1, \ldots, n,$$

as parameter functions in Groves' scheme. Then

$$\sum_i t_i(y^i) = \sum_i h_i(y^i) + (n-1) \cdot \sum_i f_i(d^*(y), y_i)$$

$$+ n \cdot c(d^*(y)) = -(n-1) \cdot p(y) + (n-1) \cdot p(y)$$

$$+ c(d^*(y)) = c(d^*(y)),$$

which is the condition for budget-balancing.

Suppose the transfer payments balance the budget. Since they are Groves' transfer payments, we have:

$$\sum_i t_i(y) = (n-1) \cdot p(y) + \sum_i h_i(y^i) - c(d^*(y)).$$

This implies, by budget-balancing (e.g., (3.19)):

$$p(y) = \frac{-1}{n-1} \sum_i h_i(y^i),$$

and consequently $p(y)$ is $(n-1)$-separable.

Q.E.D.

A corollary of the theorem (observed by Groves and Loeb [1975] in footnote 8 on p. 219) is that budget-balancing can be achieved if $p(y)$ has degree less than or equal to $(n-1)$, since then $p(y)$ must consist of terms which can contain at most $(n-1)$ of the y_i's.

In the 0-1 public project problem (Example 3.1), we have:

$$p(y) = \sum_i y_i + c, \qquad \text{if } \sum_i y_i > c,$$

$$= 0 \quad , \qquad \text{if } \sum_i y_i \leq c.$$

This function is not $(n-1)$-separable, and consequently budget-balancing cannot be achieved.[7]

Assume now that we have a problem where budget-balancing can be achieved when using Groves' scheme. It follows by (3.13) and (3.19) that:

$$y_i = \underset{m_i}{\operatorname{argmax}} \sum_j \{f_j(d^*(m_i, y^i), y_j) + t_j(m_i, y^i)\}.$$

Using our differentiability assumption, this implies

$$\frac{\partial}{\partial m_i} \sum_j \{f_j(d^*(y), y_j) + t_j(y)\} = 0,$$

$$\frac{\partial}{\partial m_i} \{f_i(d^*(y), y_j) + t_i(y)\} = 0,$$

from which we conclude that:

$$\sum_{j \neq i} \{f_j(d^*(m_i, y^i), y_j) + t_j(m_i, y^i)\} = \text{constant}, \quad \forall y^i, \forall i.$$

In other words, when budget-balancing transfer payments exist and are

used, then the other agents' summed welfare is independent of one agent's message. Each agent alone carries the social cost of changing the decision pair (d,t) via his message. In this sense, the agents are decoupled, and it is exactly this kind of decoupling that is needed to achieve full efficiency (both w.r.t. d and the transfer t_i). We will meet the same condition when we discuss efficient outcome functions in more general models.

Even though budget-balancing cannot be achieved in all circumstances with Groves' scheme, it is always possible to guarantee that the principal is left with a surplus. Assume for simplicity $c(d) \equiv 0$. Then the following scheme will do:

$$h_i(y^i) = -\max_{d \in D} \sum_{j \neq i} f_j(d, y_j), \qquad i = 1, \ldots, n.$$

This is called the pivot scheme (see Green and Laffont [1978], Loeb [1975]). With this scheme each agent pays for the externality he causes to the organization by changing the decision with his message.

Evidently, the larger the organization gets, the smaller will the expected payment by each agent become, since his message will influence less the joint decision. For example, in the 0-1 public project problem, suppose each agent's willingness to pay can be assumed to lie in a finite interval and let the agents' preferences be independent. Then the probability of anybody being a pivot, i.e., changing the social decision with his message, will go to zero as the number of agents increases. In this limiting sense the pivot scheme

seems a quite satisfactory approximate solution to the budget-balancing problem when the decision is discrete. With a continuous decision variable the individual payments will also be small in expectation, but the number of agents may make the total expected payment large, though we have not studied the question closely.

Another way to alleviate the problem of balancing the budget is to study it in a multiperiod setting. Each period may generate certain surpluses and deficits, but with some reserves such fluctuations can be accepted. What matters is the expected outcome over a longer time period. Groves [1974] has shown that if the probability distributions of agents' characteristics are independent and identical, then there exists a Groves' scheme for which the budget will balance on average. Notice that one cannot make adjustments in the scheme based on periodical deficits or surpluses, because this will destroy incentive compatibility (unless agents are assumed to behave myopically, which may be a fair assumption in many cases).

Budget-balancing can also be resolved by weakening the notion of incentive compatibility. One approach will be more thoroughly discussed in connection with the work of d'Aspremont and Gerard-Varet [1975] in Section 3.3. Here we want to mention the work of Hurwicz [1976] and Groves and Ledyard [1977]. Both have been successful in constructing budget-balancing schemes, which require an iterative process for achieving an optimal solution. These schemes are incentive compatible in the weaker sense that telling the truth is a Nash equilibrium in the game with messages rather than message functions

as strategies. In other words, strategic behavior in the iterative game is not analyzed. The reason budget-balancing can be achieved, is that the dimensionality of the messages can be reduced in an iterative process, and this will make the social net deficit function in (3.20) become (n - 1)-separables (compare to the result that if p(y) is polynomial with degree less than or equal to (n - 1), then budget-balancing is possible).

3.2.4 Individual Rationality

Another shortcoming of Groves' scheme is that the outcome function it generates will not be individually rational for all y, unless we allow a budget deficit. This means that we cannot guarantee that a Pareto move is made when Groves' scheme is used. For instance, look again at the 0-1 public project problem. Individual rationality requires:

$$y_i + \sum_{j \neq i} m_j + h_i(m^i) \geq 0, \quad \text{if} \quad \sum_j m_j > c,$$

$$h_i(m^i) \geq 0, \quad \text{if} \quad \sum_j m_j \leq c.$$

For any m^i, there exists an m_i such that $\Sigma\, m_j < c$; hence, $h_i(m^i) \geq 0$, $\forall m^i$, $\forall i$. This implies a budget deficit, whenever the project is undertaken.

Groves and Loeb [1975] have tried to get around the problem of individual rationality in the case of a continuous public decision in the following way. They design an incentive compatible scheme, which

parameterized by "cost shares" $\theta = (\theta_1, \ldots, \theta_n)$, $\sum_i \theta_i = 1$, and has the property that for any θ in the unit simplex, a budget surplus is guaranteed. Next they show that there exists a set of θ-values for which the scheme is individually rational. The size of this set depends on how far one is from Pareto optimality currently, but it is not possible to name the individually rational θ-values before one knows the exact preference profiles of the agents (which is the information one is looking for in the first place).

For this reason, they argue, one should let the agents bargain about what θ-value to use. Once they have reached an agreement, an efficient outcome is guaranteed using Groves' scheme. According to Groves and Loeb such a bargaining process is a simple way of finding a satisfactory θ-value. We cannot quite agree with this statement. Why would it be easier to bargain about cost shares than about the joint decision directly? Notice that in bargaining over his cost share, an agent does not know what his total cost will be, since this depends on the other agents' preferences. If agents have subjective beliefs about other agents' preferences and use these as a basis in their bargaining, then it is perfectly possible that the status quo is an efficient point in such a game.

Because individual rationality cannot be achieved in the strong sense described above, a weaker notion has been proposed (see Thomson [1976]). A scheme is said to be <u>weakly individually rational</u> if each agent prefers to participate in the revelation game rather than stay out. That is, once it has been decided that a certain decentralized

decision mechanism will be used, everybody wants to participate. Weak incentive compatibility can be achieved by using the pivot scheme. In fact, it is the unique scheme which is both weakly individually rational and feasible (never gives a budget deficit) for all values of y. This makes the pivot scheme look quite desirable. However, we observe that using the pivot scheme may lead to a Pareto inferior state. For instance, in the public project problem, whenever the project is rejected and somebody is a pivot, this is the case.

We will return to a discussion of individual rationality in the next section.

3.3 Weak Incentive Compatibility

3.3.1 Partial Communication

In the preceding model it was assumed that agents can communicate their full information y_i to the principal. This assumption is, of course, fundamental for achieving efficient decisions in each information state y. We now turn to the case in which only partial communication is possible. This is the framework in which Groves originally discovered his incentive compatible mechanism (see Groves [1973]). We will not be as general as Groves, since we will assume that messages are finite dimensional, and communication takes place only from agents to the principal. Our main purpose is to show that the uniqueness result, appropriately generalized, will still be true.

When communication is restricted, the situation is rather different from before. We have to view the problem in the framework

of a game of incomplete information. The reader is referred to Chapter I for a general description of such a game formulation. The objective of the principal is to induce the agents to communicate their information optimally. It will become clear shortly that one cannot hope for dominant message strategies, so the appropriate notion of incentive compatibility is based on a standard Nash-equilibrium requirement.

Let y_i, $i = 1, \ldots, n$, be the agents' information signals and $f_i(d,y_i)$ their preference functions. We assume that each agent's message m_i belongs to $\mathbb{R}^{k_i}$. The social optimum entails optimal message strategies $\{m_i^*(y_i)\}$, and an optimal decision function $d^*(m)$. Notice that $m_i^*(y_i) = y_i$ is no longer feasible, since it is assumed that y_i has higher dimension than m_i. From the theory of teams (see Marschak and Radner [1972]) we know that:

$$(3.21) \qquad (m^*(y),d^*(m)) = \underset{m(\cdot),d(\cdot)}{\operatorname{argmax}} \; E\{\sum_{i=1}^{n} f_i(d(m(y)),y_i)\}.$$

The question is: how should the compensation rules $t_i(m)$ be chosen, so that $\{m_i^*(y_i)\}$ will be a Nash equilibrium in the game of incomplete information. The answer is given in Groves [1973], under the assumption that the y_i's are independent; choose:

$$(3.22) \qquad t_i(m) = g_i(m;h_i) \equiv E\{\sum_{j\neq i} f_j(d^*(m),y_j) \,|\, m^*(y) = m\} + h_i(m^i), \; \forall i.$$

This is the natural extension of (3.15). Because of the independence

assumption, (3.22) can be written:

$$(3.23)\quad g_i(m;h_i) = \sum_{j\neq i} E\{f_j(d^*(m),y_j)\,|\,m_j^*(y_j) = m_j\} + h_i(m^i).$$

In fact, the step from (3.22) to (3.23) could be taken under the weaker assumption that y_i and $m^{i*}(y^i)$ are conditionally independent, given $m_i^*(y_i)$. This would also suffice for proving the incentive compatibility of Groves' scheme, but we will maintain Groves' assumption for a simpler uniqueness characterization.[8]

Our main theorem is the following:

Theorem 3.3: A transfer scheme $t_i(m)$, $i = 1, \ldots, n$, is incentive compatible under partial communication with independent observations y_i, $i = 1, \ldots, n$, if and only if it equals a Groves' scheme in expectation, that is,

$$(3.24)\qquad E\{t_i(m_i,m^{i*}(y^i))\} = E\{g_i[(m_i,m^{i*}(y^i));h_i]\},\ \forall m_i,\ \forall i,$$

for some choice of h_i-functions in (3.23).

Before proving the theorem let us define some notation. We will write $\tilde{m}_i^*$ for the random variable $m_i^*(y_i)$, and m_i^* for its outcomes. Define the functions:

$$\bar{f}_i(d,m_i^*) = E\{f_i(d,y_i)\,|\,\tilde{m}_i^* = m_i^*\},\ \forall i.$$

By (3.21) we have:

$$(3.25) \qquad d^*(m^*) = \underset{d}{\operatorname{argmax}} \sum_{i=1}^{n} \bar{f}_i(d, m_i^*).$$

This implies:

$$(3.26) \qquad m_i^* = \underset{m_i}{\operatorname{argmax}} \sum_{j=1}^{n} \bar{f}_j(d^*(m_i, m^{i*}), m_j^*), \qquad \forall m^*, \ \forall i.$$

As before, we need a differentiability assumption. We will assume that the functions:

$$(3.27) \quad S_i(m_i; m_i^*) \equiv E\{\sum_{j=1}^{n} \bar{f}_j(d^*(m_i, \tilde{m}^{i*}), \tilde{m}_j^*) \,|\, \tilde{m}_i^* = m_i^*\}, \ i = 1, \ldots, n,$$

are differentiable w.r.t. m_i at m_i^*, for every m_i^*.

Proof of Theorem 3.3:

Sufficiency: Suppose the transfer scheme satisfies (3.24). The agent's problem is to maximize:

$$E\{f_i(d^*(m_i(y_i), \tilde{m}^{i*}), y_i) + t_i(m_i(y_i), \tilde{m}^{i*})\},$$

w.r.t. the function $m_i(y_i)$; or by (3.24), to maximize

$$(3.28) \quad E\{f_i(d^*(m_i(y_i), \tilde{m}^{i*}), y_i) + \sum_{j \neq i} \bar{f}_j(d^*(m_i(y_i), \tilde{m}^{i*}), \tilde{m}_j^*) + h_i(\tilde{m}^{i*})\}.$$

The function h_i will play no role in this maximization so we can ignore it.

From (3.21) we know that $m_i^*(y_i)$ maximizes:

$$E\{\sum_{j=1}^{n} f_j(d^*(m_i(y_i),\tilde{m}^{i*}),y_j)\}.$$

This can be written:

$$(3.29) \qquad E[f_i(d^*(m_i(y_i),\tilde{m}^{i*}),y_i)] + \sum_{j\neq i} E[f_j(d^*(m_i(y_i),\tilde{m}^{i*}),y_j)].$$

Comparing (3.28) and (3.29), it suffices to show that,

$$E[\bar{f}_j(d^*(m_i(y_i),\tilde{m}^{i*}),\tilde{m}_j^*)] = E[f_j(d^*(m_i(y_i),\tilde{m}^{i*}),y_j], \quad \forall j\neq i.$$

But this equality is immediate, because the expectation on the left hand side is an iterated expectation, which reduces to the expression on the right.

Necessity: Suppose $\{t_i(m)\}$ is a set of incentive compatible transfer payments. Define

$$h_i(m) = t_i(m) - \sum_{j\neq i} \bar{f}_j(d^*(m),m_j).$$

Let $Y_i(m_i^*) = \{y_i | m_i^*(y_i) = m_i^*\}$. From the agent's optimality condition we have that:

$$m_i^* = \underset{m_i}{\text{argmax}}\ E_{y^i}\{f_i(d^*(m_i,\tilde{m}^{i*}),y_i) + \sum_{j\neq i} \bar{f}_j(d^*(m_i,\tilde{m}^{i*}),\tilde{m}_j^*)$$

$$+ h_i(m_i,\tilde{m}^{i*})\},\quad \forall y_i \varepsilon Y_i(m_i^*).$$

Integration over $Y_i(m_i^*)$ yields:

$$(3.30)\qquad m_i^* = \underset{m_i}{\text{argmax}}\ \{E\{\sum_{j=1}^{n} \bar{f}_j(d^*(m_i,\tilde{m}^{i*}),\tilde{m}_j^*)\,|\,\tilde{m}_i^* = m_i^*\}$$

$$+ E\{h_i(m_i,\tilde{m}^{i*})\}\}.$$

From (3.26) follows, by integration over $Y_i(m_i^*)$:

$$(3.31)\qquad m_i^* = \underset{m_i}{\text{argmax}}\ E\{\sum_{j=1}^{n} \bar{f}_j(d^*(m_i,\tilde{m}^{i*}),\tilde{m}_j^*)\,|\,\tilde{m}_i^* = m_i^*\}.$$

It is then clear from (3.30), (3.31) and the differentiability assumption (3.27), that Lemma 3.1 applies, and we can conclude that:

$$E\{h_i(m_i,\ \tilde{m}^{i*})\} = \text{constant}.$$

In view of the definition of h_i, this statement is equivalent to (3.24). This concludes the proof.

Q.E.D.

The first part of the proof is exactly the same as in Groves [1973], with appropriate simplifications due to our particular assumptions. The second part is quite similar to the proof of Theorem 3.2, except for some complications that arise because the principal and the agent have different information even after the agent has sent his message.

It is evident that we cannot generally expect the agent to have a dominant strategy. This would require that:

$$m_i^* = \underset{m_i}{\text{argmax}} \{f_i(d^*(m_i,m^i),y_i) + \sum_{j\neq i} \bar{f}_j(d^*(m_i,m^i),m_j)\},$$

$$\forall m^i,\ \forall y_i \varepsilon Y_i(m_i^*).$$

But in that case m_i^* would provide all the relevant information about y_i, which is generally false, since we have partial communication.

Secondly, we observe again that if the assumption of conditional independence is not satisfied, then Groves' scheme will not be incentive compatible. In the proof, (3.28) would not be correct and the agent's and the principal's preferences would be different. We can illustrate this with a simple example.

Example 3.2: There are two divisions and a center. The profit functions are:

$$f_1(d,y_1) = y_1 \cdot d,$$

$$f_2(d,y_2) = (y_2 - 900) \cdot d.$$

$d = 0$ or 1. y_1 and y_2 are perfectly correlated random variables, which can take on values $y_1 = 0$ or 1, $y_2 = 0$ or 1000. $P(y_2 = 0|y_1 = 0) = 1$, $P(y_2 = 1000|y_1 = 1) = 1$. Only division 1 communicates with the center. It is supposed to report the true value of y_1.

Obviously, $d^*(m_1) = 0$, if $m_1 = 0$, and $d^*(m_1) = 1$, if $m_1 = 1$. However, with Groves' scheme division 1 will always report $y_1 = 1$ to get a profit of either 100 or 101, depending on the true value of y_i.

The point is that with dependence, the agents will not only affect their payments indirectly via $d^*(m)$, but also directly by changing the expectation. This will make the principal's and the agent's preference functions incompatible. □

One of the important implications of Theorem 3.4 is that it shows that Groves' scheme is essentially unique even if only approximations to the true preference functions are communicated. It might have been tempting to make this conclusion already after Theorem 3.2, since a natural interpretation of the finite-dimensional parameters is that they are coefficients of approximations to the true preference functions (see for instance Loeb [1975] for an interpretation of this kind). However, we could not do that, because with incomplete

communication we loose strong incentive compatibility, and Theorem 3.2 no longer applies.

This is important to keep in mind. It may be that the procedures developed in the previous section for strong incentive compatibility lead to very small errors when most of the information can be communicated, but from a theoretical standpoint, partial communication can only be treated satisfactorily in a game of incomplete information, as has been done in this section.

3.3.2 Full Communication

Partial communication forced us to model the problem as a game of incomplete information. We will now study full communication in the same framework, even though dominant strategies would be available. The reason is that with the weaker notion of incentive compatibility we can achieve a balanced budget. This idea is due to d'Aspremont and Gerard-Varet [1975], and we will follow their presentation closely.

Since Groves' scheme is strongly incentive compatible, we get the following sufficient condition:

Theorem 3.4 (d'Aspremont and Gerard-Varet): A transfer scheme $t_i(m)$, $i = 1, \ldots, n$, which satisfies:

$$(3.32)\quad E\{t_i(m_i,y^i)\,|\,y_i\} = E\{g_i(m_i,y^i;h_i)\,|\,y_i\}, \qquad \forall m_i,\ \forall y_i,\ \forall i,$$

for some choice of h_i-functions, is weakly incentive compatible.

Proof: Agent i will maximize:

$$E\{f_i(d^*(m_i,y^i),y_i) + t_i(m_i,y^i)|y_i\}$$

$$= E\{f_i(d^*(m_i,y^i),y_i) + g_i(m_i,y^i)|y_i\}, \quad \forall y_i,$$

where d^* is defined by (3.6). Since $m_i = y_i$ is a dominant strategy when Groves' scheme is used, the integrand is pointwise maximized by $m_i = y_i$, and hence truth-telling will be optimal.

Q.E.D.

To see that (3.32) is not generally a necessary condition, we can look at the following example.

Example 3.3: There are two agents. y_1 and y_2 are distributed so that conditional on y_i, y_{i+1} has a normal distribution with mean y_i and variance σ_i^2 (addition modulo 2). Then the following parameter functions h_i in the Groves' scheme will work:

$$h_i(m) = -(m_i - m_{i+1})^2, \quad i = 1, 2.$$

To see this, note that

$$(3.33) \qquad E(h_1(m_1,y_2)|y_1) = -\sigma_2^2 - (m_1 - y_1)^2.$$

Hence,

$$E(f_1(d^*(m_1,y_2),y_1) + f_2(d^*(m_1,y_2),y_2) - (m_1 - y_2)^2|y_1)$$

will be maximized at $m_1 = y_1$, so the scheme is incentive compatible. However, it does not satisfy (3.32), which is immediate from (3.33).

□

If we assume that observations are independent, and that the functions:

$$S_i(m_i,y_i) = E\{\sum_{j=1}^{n} f_j(d^*(m_i,y^i),y_j)|y_i\}, \quad i = 1, \ldots, n,$$

are differentiable, then it follows directly from Theorem 3.3 that (3.32) is also necessary.

Theorem 3.5 (d'Aspremont and Gerard-Varet): A transfer scheme $t_i(m)$, $i = 1, \ldots, n$ which is weakly incentive compatible with independent observations, satisfies

$$E\{t_i(m_i,y^i)\} = E\{g_i(m_i,y^i;h_i)\}, \quad \forall m_i, \forall i, \tag{3.34}$$

for some choice of h_i-functions in (3.23).

Remark: For the case of dependent observations, we have only been able to develop the partial differential equation condition:

$$(3.35)\quad \frac{\partial}{\partial m_i} E\{t_i(y_i,y^i)\,|\,y_i\} = \frac{\partial}{\partial m_i} E\{g_i(y_i,y^i;h_i)\,|\,y_i\}, \quad \forall y_i,\ \forall i.$$

With the independence assumption, it is easy to achieve budget-balancing in addition to incentive compatibility. For instance, proceed as follows. Start with Groves' scheme, taking $h_i \equiv 0$, $\forall i$. This leads to a social deficit

$$(3.36)\quad p(m) = -(n-1) \sum_{i=1}^{n} f_i(d^*(m),m_i), \quad \forall n.$$

Allocate this function arbitrarily between the agents, so that if $p_i(m)$ is agent i's share, then

$$p(m) = \sum_{i=1}^{n} p_i(m), \quad \forall m.$$

Define,

$$\bar{p}_i(m_i) = E_{y^i}[p_i(m_i,y^i)], \quad \forall i.$$

Allocate each $\bar{p}_i(m_i)$ among agents, excluding the i^{th} agent; e.g., give each agent $j \neq i$ an equal share of $\bar{p}_i(m_i)$.

As a result of the construction we get the following transfer payments:

$$(3.37)\quad t_i(m) = g_i(m;0) + p_i(m) - \bar{p}_i(m_i) + \frac{1}{n-1}\sum_{j\neq i}\bar{p}_j(m_j),\ i = 1, \ldots, n.$$

Taking the expectation of t_i we find that:

$$E(t_i(m_i,y^i)) = E(g_i(m_i,y^i;h_i)),$$

where $h_i(y^i) = \frac{1}{n-1}\sum_{j\neq i}\bar{p}_j(y_j)$. Hence, (3.37) defines an incentive compatible transfer scheme by Theorem 3.4.

We can see that there is substantial freedom in choosing among budget-balancing and incentive compatible transfer schemes, since both $p(m)$ and $\bar{p}_i(m_i)$ could be allocated arbitrarily, and we could also have started with an arbitrary set of h_i-functions. A natural question then is: what other desirable properties can we achieve by a proper choice of transfer payments? In particular, can we achieve individual rationality?

This question has been studied by Kobayashi [1977] in the context of a fixed-sized public project problem. Kobayashi shows that one can choose the functions t_i such that no agent pays anything if the project is rejected; hence, nobody will be worse off in that event. However, with such a choice of transfer payments there may be individuals who will be worse off than before when the project is accepted. In this sense only partial individual rationality can be guaranteed. Kobayashi shows further that a weaker notion of individual rationality can be achieved; namely, the functions t_i can be chosen so

that each agent is better off in expectation (ex ante) when the scheme is implemented and agents report honestly.

It is clear that the preceding construction of budget-balancing schemes cannot be carried out when observations are dependent. The reason is that the $\bar{p}_i$-functions will then depend on y_i, which in turn disturbs the other agents' incentives when reallocated as in (3.37). Of course, it would suffice to have one agent whose observation is independent of the other agents' observations, in order to achieve budget-balancing. It is an open question whether this can be done when all observations are mutually dependent.

3.4 Extensions

What is driving the possibility results from the previous sections? Apparently the fact that agents' preference functions can be transformed to become equivalent to the social welfare function. This again rests on the assumption that agents' preferences are additively separable and linear in money. But linearity is not essential in itself; only the fact that with linear functions we know precisely the effects of transfer payments. We could, of course, achieve the same results with nonlinear utilities over money, provided the form of the nonlinearity would be known to the principal. On the other hand, separability seems fundamental for achieving a possibility result. Green and Laffont [1977] give an example, which shows that for a certain type of nonseparable utility functions, strong incentive

compatibility can never be achieved.

It is of interest to find more general conditions under which incentive compatibility can be achieved. In this section we explore some relaxations of our earlier assumptions. First, we show that some problems, which do not directly fit the assumptions on preferences that we have made, can be transformed to be applicable to Groves' scheme. Secondly, we look at two examples where agents' observations are dependent; one of which works, and the other one not. Finally, we show an impossibility result related to two-person exchange economies.

3.4.1 Revelation of Risk-Tolerances in a Syndicate

There are n agents, each with an exponential utility function over wealth:

$$u_i(x) = -e^{-\frac{1}{\rho_i} \cdot x}, \quad i = 1, \ldots, n.$$

ρ_i, the risk-tolerance of agent i (see Wilson [1968]), is only known to agent i. As a syndicate they face the problem of making a decision $\underline{a}$ and sharing the outcome $x(a,z)$, where z is a random variable with a distribution known to each agent.

The question is: can we design a decision mechanism, which specifies what action $\underline{\hat{a}}$ should be taking and what sharing rules $s_i(x)$, $i = 1, \ldots, n$ should be employed, such that it leads to an efficient outcome for all preference profiles $\rho = (\rho_1, \ldots, \rho_n)$? The answer is in the affirmative if we do not require that the budget balances.

The form of the utility functions are such that at first sight the task may not appear possible. A simple transformation of the problem shows, however, that we face essentially the same situation as in the earlier sections. Instead of writing the social objective function as a weighted sum of expected utilities, we can equivalently write it as:[9]

$$\sum_{i=1}^{n} \lambda_i \cdot c_i(a,s_i), \qquad \lambda_i > 0, \ \forall i. \tag{3.38}$$

Here $c_i(a,s_i)$ stands for the certainty equivalent of agent i when a decision $\underline{a}$ is taken and he gets a share $s_i(x)$ in the outcome. Maximizing (3.38) over $\underline{a}$ and $\{s_i\}$ will produce a Pareto-optimal action and sharing rule.

With exponential utility functions we know that the certainty equivalent is additively separable in transfer payments, i.e.,

$$c(a,s_i + t_i;\rho_i) = c(a,s_i;\rho_i) + t_i, \tag{3.39}$$

where t_i is a constant, and $c(\cdot,\cdot;\rho_i)$ denotes the certainty equivalent for an exponential utility function with risk-tolerance ρ_i. Because of the form of (3.39), the efficient decision and sharing rule satisfies:

$$(a^*(\rho),\{s_i^*(\rho)\}) = \operatorname{argmax} \sum_{i=1} c_i(a,s_i;\rho_i). \tag{3.40}$$

Since the sum of the certainty equivalents of exponential

utility functions is a certainty equivalent for a surrogate utility function, (3.40) states the well-known result that the syndicate's behavior can be characterized by a utility function with risk-tolerance equal to the sum of the agents' risk-tolerances (see Wilson [1968]).

From (3.40) we can immediately see the following result:

Proposition 3.6: In the case of exponential utility functions, a s.i.c. transfer scheme $t_i(r)$, $i = 1, \ldots, n$, takes the form:

$$(3.41) \qquad t_i(r) = \sum_{j \neq i} c(a^*(r), s_j^*(r); rj) + h_j(r^i), \quad \forall i,$$

where h_j can be chosen arbitrary, but independent of r_i.

Proof: Follows from (3.40) and Theorem 3.2.

Q.E.D.

The efficient action function $a^*(\rho)$ depends, of course, on the particular outcome function $x(a,z)$ and on the distribution of z. The efficient sharing rule, on the other hand, is according to the syndicate theory always of the form:

$$(3.42) \qquad s_i^*(x;\rho) = \frac{\rho_i}{\bar{\rho}} \cdot x + k_i(\rho), \quad \forall i; \ \sum_i \rho_i = \bar{\rho}.$$

Here $k_i(\rho)$ may be arbitrary functions which sum to zero.

As an illustration of the proposition we can look at an

example where the syndicate is only involved in sharing a normally distributed risk $\tilde{x}$. Let the mean of $\tilde{x}$ be μ and the variance σ^2. The incentive compatible transfer scheme becomes (using (3.42)):

$$t_i(\rho) = \frac{\bar{\rho}^i}{\bar{\rho}} [\mu - \frac{1}{2} \cdot \frac{1}{\bar{\rho}} \cdot \sigma^2] + h_i(\rho^i), \ \forall i, \tag{3.43}$$

where $\bar{\rho}^i = \sum_{j \neq i} \rho_i$. The term in the bracket is the syndicate's certainty equivalent. Since the agent's certainty equivalent with s_i^* (letting $k_i(\rho) = 0$) is:

$$\frac{\rho_i}{\bar{\rho}} \cdot \mu - \frac{1}{2} \cdot \frac{1}{\rho_i} \cdot \frac{\rho_i^2}{\bar{\rho}^2} \cdot \sigma^2 = \frac{\rho_i}{\bar{\rho}} [\mu - \frac{1}{2} \cdot \frac{1}{\bar{\rho}} \cdot \sigma^2],$$

the transfer payment in (3.43) imposes the syndicate's certainty equivalent on the agent, and everybody will act in the best interest of the syndicate as a whole.

We notice that with the transfer payments in (3.43) the budget cannot be balanced, because the net deficit function (defined in (3.20)) is not (n-1)-separable.

3.4.2 Dependent Observations

We argued earlier that Groves' scheme cannot be used when observations are dependent, and illustrated the point in Example 3.2. We will first show that in the syndicate of the previous section no scheme can achieve incentive compatibility when agents have some private information about the distribution of $\tilde{x}$. Secondly, we give

an example with dependence for which one can find an incentive compatible scheme (but not using a Groves' scheme, of course).

Look again at the situation where the syndicate only shares a risk $\tilde{x}$. Assume now that instead of having homogeneous beliefs about $\tilde{x}$, each agent has observed a private signal, say a random sample about $\tilde{x}$, and bases his probability beliefs on this signal accordingly. From the syndicate theory we know that the efficient shares do not depend on these signals. Consequently, the decision mechanism must be of the form:

$$(3.44) \qquad s_i(x;m,r) = \frac{r_i}{\bar{r}} \cdot x + k_i(m,r), \quad i = 1, \ldots, n,$$

where $m = (m_1, \ldots, m_n)$ are the messages about the signals $y = (y_1, \ldots, y_n)$, and r are the messages about the risk-tolerances. But clearly $k_i(m,r)$ cannot depend on m_i, since the agent would just pick the message m_i which gave him the highest transfer. Hence, his outcome is independent of m_i. In that case it is evident that he will distort his risk-tolerance message depending on his signal about x, in order to get a larger share if the signal is favorable and a smaller share if it is less favorable.

We conclude that the additional sample observations will destroy incentive compatibility. The result is a variation of the well-known theme that additional information destroys insurance opportunities.

Let us turn now to a numerical example where it is possible to achieve incentive compatibility.

Example 3.3: There are two divisions with profit functions:

$$f_1(x,z_1) = z_1 \cdot x,$$

$$f_2(x,z_2) = -z_2 \cdot x - \frac{1}{2} \cdot x^2.$$

$z = (z_1,z_2)$ is the state of nature. Divisions observe signals y_1 and y_2 respectively. The probability structure is such that:

$$E(z_1|y_1,y_2) = .9y_1 + .1y_2,$$

$$E(z_2|y_1,y_2) = .1y_1 + .9y_2.$$

A straightforward calculation shows that the scheme:

$$t_1(m) = -(.36\ m_1^2 + .08\ m_1 \cdot m_2),$$

$$t_2(m) = -(.04\ m_2^2 + .72\ m_1 \cdot m_2),$$

makes truth-telling a best response against any true signal of the other agent. We underline the word true, because we do not have dominant strategies, of course. However, we do have a Nash equilibrium independently of the distributions of y_1 and y_2, which in some sense could be regarded as a semi-strong form of incentive compatibility.

□

When one tries to study more generally the case of dependent observations, one runs into the problem that, even though first-order optimality conditions can be easily guaranteed for truth-telling strategies, it seems hard to determine the global optimality of such strategies. We have not pursued the topic further.

3.4.3 A Two-Person Exchange Economy

For a decision function $d^*(y)$ to be s.i.c. we must have:

$$(3.45) \qquad y_i = \underset{m_i}{\operatorname{argmax}}\ f_i(d^*(m_i, y^i), y_i), \qquad \forall y, \forall i.$$

For $d^*(y)$ to be efficient at every y it must satisfy (assuming a convex Pareto frontier):

$$(3.46) \qquad d^*(y) = \underset{d}{\operatorname{argmax}} \sum_{j=1}^{n} \lambda_j(y) f_j(d, y_j),$$

for some $\lambda_j(y)$-functions which are strictly positive. From (3.46) we derive the weaker condition:

$$(3.47) \qquad y_i = \underset{m_i}{\operatorname{argmax}} \sum_{j=1}^{n} \lambda_j(y) f_j(d^*(m_i, y^i), y_j), \qquad \forall y, \forall i.$$

Our standard technique has been to combine (3.45) and (3.47) for some implications about d^*. We will use it again in the context of a two-person exchange economy.

Let there be two agents and ℓ commodities. The total amount

of resources are assumed known and fixed, and the problem is to allocate them efficiently between the two agents. The utility functions of the agents are unknown, but assumed to satisfy the standard convexity and continuity assumptions. As before, a decision mechanism is employed, which asks for the agents' preferences and takes an efficient action if the agents tell the truth. Such mechanisms could be iterative, the price mechanism being one example.

Hurwicz [1972] asked whether we can find a mechanism for which each agent would tell the truth as a dominant strategy. He showed first that the price mechanism is not s.i.c. in general with a finite number of agents, and then extended the argument to show that no mechanism which is individually rational could be s.i.c. in a sufficiently rich domain of utility functions. Hurwicz's argument is based on the existence of monopolistic prices and is quite different from ours. We will here derive the impossibility result directly from the two conditions (3.45) and (3.47) without assuming individual rationality.

Let $d_0 = (d_1,d_2)$ be an interior allocation, such that it is efficient for a pair of utility functions $f_1(d_1)$, $f_2(d_2)$. Let $f_1(d_1;y_1)$, $f_2(d_2;y_2)$ be two families of utility functions, parametrized by $y_1,y_2 \in \mathbb{R}^{2\ell}$, such that $f_1(d_1) = f_1(d_1;0)$, $f_2(d_2) = f_2(d_2;0)$. We will impose further conditions on this parametrization later; now it is taken such that f_1 and f_2 are smooth w.r.t. the parameters.

Let $d^*(y)$ represent a mapping from the parameter space $\mathbb{R}^{2\ell}$ to efficient allocations for the corresponding utility functions,

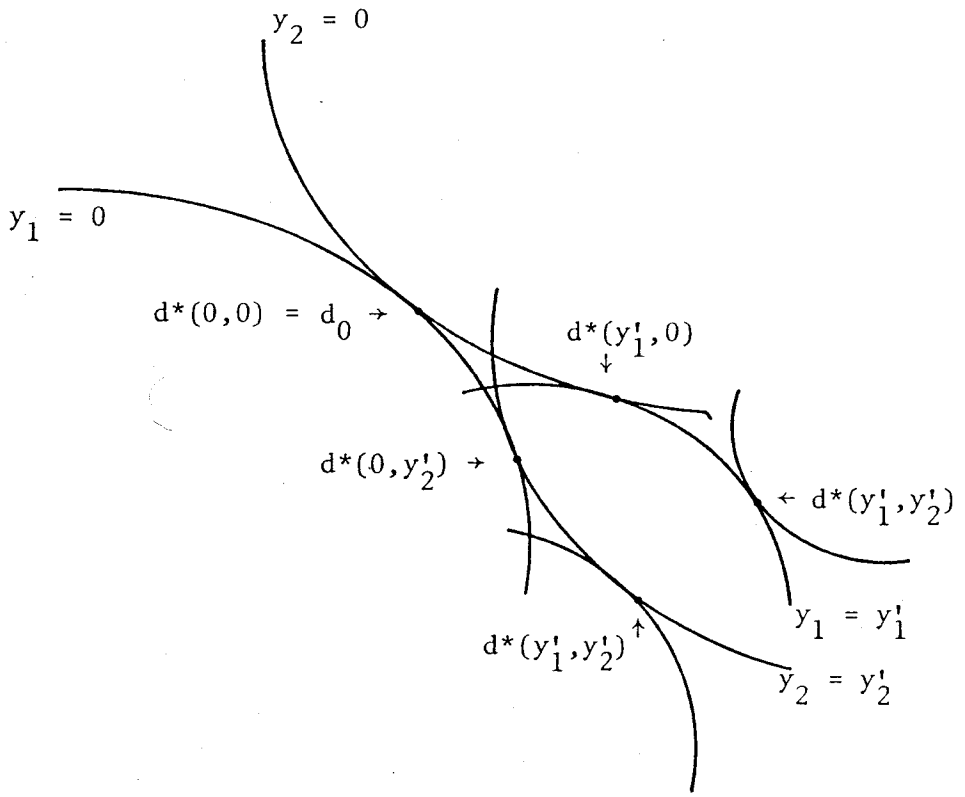
which satisfies $d^*(0,0) = d_0$. We want to show that the parametrization can be chosen so that $d^*(y)$ cannot be attained via dominant strategies.

Because of the smoothness of f_1 w.r.t. y_1, $f_1(d_1^*(m_1,y_2);y_1)$ must be differentiable w.r.t. m_1 in order for (3.45) to hold true (we omit the proof, which is similar to Lemma 3.1). $f_2(d_2^*(m_1,y_2);y_2)$ is independent of y_1. Consequently, using the fact that,

$$(3.48)\quad y_1 = \underset{m_1}{\operatorname{argmax}} \{\lambda(y) f_1(d_1^*(m_1,y_2);y_1) + f_2(d_2^*(m_1,y_2);y_2)\},$$

for some $\lambda(y)$-function, we can conclude from Lemma 3.1 that $f_2(d_2^*(m_1y_2);y_2)$ is independent of m_1. Likewise, $f_1(d_1^*(y_1,m_2);y_1)$ is independent of m_2.

Let $D_1(y_1) = \{d \mid d = d^*(y_1,y_2)$ for some $y_2\}$, and define correspondingly $D_2(y_2)$. The argument above shows that $D_1(y_1)$ and $D_2(y_2)$ are subsets of indifference surfaces of $f_1(d_1;y_2)$ and $f_2(d_2;y_1)$ respectively; (by an appropriate choice of the parametrization we can, indeed, get them to coincide with these indifference surfaces). From this it is easy to show that $d^*(y)$ must be overdetermined. Study the picture below:



Starting from $d^*(0,0)$, we determine $d^*(y_1',y_2')$ in two ways, keeping in mind that $D_1(y_1)$ and $D_2(y_2)$ are indifference curves. $d^*(y_1',0)$ has to lie on the indifference curve $D_2(0)$ through d_0. $d^*(y_1',y_2')$ has to lie on the indifference curve $D_1(y_1')$ through $d^*(y_1',0)$. On the other hand, $d^*(y_1',y_2')$ has to lie on the indifference curve $D_2(y_2')$ through $d^*(y_2',0)$. Because $d^*(y_1',y_2')$ is efficient, this must imply that $d^*(y_1',0) = d^*(0,y_2') = d_0$. But this is not, of course, possible for all y_1',y_2'.

We have thus informally shown:

Proposition 3.8: In a two-person exchange economy, efficient outcome functions cannot be attained via dominant strategies.

From the construction above we see that budget-balancing is sufficient to destroy incentive compatibility in the exchange economy. Individual rationality is not needed. The preceding argument does not extend directly to an n-person economy, and we will not pursue the issue further here.

3.5 Conclusions

In this chapter we have studied coordination of information mainly in the context of additively separable preference functions. The objective has been to analyze under what conditions efficient outcome functions can be attained. Two tools have been used in this analysis. First, we noticed that one can check the attainability of an outcome function by using it as a decision function and see whether agents will report their information honestly. Secondly, the efficiency condition could be written in terms of a maximization over each agent's message. This enabled us to combine the Nash equilibrium with the efficiency condition to infer properties of the outcome function. We found that Groves' scheme and the uniqueness thereof, appeared naturally from this construction. Moreover, the same technique provided a new uniqueness result in the case of incomplete communication.

We think that our methodology is insightful for understanding when efficiency can be achieved with a decentralized decision mechanism.

It is apparent that this is only possible in quite restricted environments, and as an illustration we exhibited an impossibility result by Hurwicz [1972], which showed that Pareto optimality cannot be guaranteed in a two-person exchange economy.

Footnotes to Chapter III

[1]The notion of efficiency we will use in this chapter is the traditional one under perfect information.

[2]Recall the discussion in Chapter I, which showed that attainability can be checked by choosing the outcome function as the decision function and see if it induces truth-telling.

The reader is also reminded of the use of superscripts;

$y^i = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$ and

$(m_i, y^i) = (y_1, \ldots, y_{i-1}, m_i, y_{i+1}, \ldots, y_n)$.

[3]This can be stated alternatively as: every Nash equilibrium in the game consists of dominant strategies.

[4]I am indebted to Takao Kobayashi for pointing out that the differentiability assumption on transfer payments can be dropped. Lemma 3.1 is due to him.

[5]For an extensive discussion of various aspects of this problem, see Green and Laffont [1978].

[6]Originally, Green and Laffont [1977] proved their uniqueness theorem for this special case.

[7]A proof of this can be found in Green and Laffont [1978].

[8]Though conditional independence rather than full independence is a minor extension of Groves' work, it is important from a practical point of view. If we think of resource allocation in a firm, it is likely that the divisions' information sets are highly dependent, for instance, due to general economic conditions, but natural that each division has incorporated this general information in their reports. This would imply conditional independence in the sense we have discussed.

[9]The equivalence follows directly from the definition of a certainty equivalent and the monotonicity of the utility functions.

CHAPTER IV

PRODUCTION INCENTIVES

4.1 Introduction

In the two previous chapters we have dealt with decentralization problems, that is, problems caused by differential information about state uncertainty. We now turn to the second major source of incentive problems, namely nonobservability of productive inputs provided by members of the organization.

Two situations will be analyzed as representative of the kinds of problems that may occur. First we look at team production. n agents jointly determine a monetary outcome by taking private actions which are nonobservable to the principal. An agent's action is costly only to himself, and the cost cannot be observed either. The problem is to induce agents to take efficient actions for optimal team production. Since it is not possible to infer from the outcome what action each individual took, an agent can cover dysfunctional behavior by blaming the others for the deviation in outcome. We want to know how agents' rewards should depend on the outcome so that proper behavior results.

We show first that if the agents form a partnership and hence have to share the outcome fully between themselves, there exists no

sharing rule based on the total outcome alone, which induces proper incentives for action. In that case a richer set of observable variables is needed. In particular, if we impose the restriction that shares should be monotone, then n independent measures of the outcome are required, which will effectively discern the actions taken by the agents.

A similar model has been studied by Kleindorfer and Sertel [1976], but rather than looking at ways of alleviating the problem of inefficiency, they study the optimal second-best solution. Alchian and Demsetz [1972] also address the question of efficiency in team production. They conclude correctly that observing only the total outcome, is insufficient for efficiency. From this they argue that competition will lead the partnership structure to fall apart and develop into an organization where there is a monitor who will control that agents take correct actions. In order to induce the monitor to perform his job properly, he should be given a residual of the outcome. This will guarantee efficiency according to Alchian and Demsetz.

This line of reasoning provides a theory of the firm. Firms develop since their organizational structure is superior to a partnership as argued above. Their analysis does not explain, however, the existence of corporations, where part of the residual goes to stock owners who do a very limited amount of monitoring themselves. We will argue that monitoring can largely be dispensed with (in the context of certainty) simply be letting outsiders, who provide no inputs for production, pick up the residual. This may be taken as one explanation

for separation of ownership from production. We also notice that adding a monitor will extend the team and improve its performance, but inefficiencies will still remain for the same reasons as earlier. This will be avoided when a separate ownership is installed.

Our second and main model is one of moral hazard in contracting. An agent is hired by a principal to provide some service. Whether or not the agent takes proper actions for providing the service cannot be observed directly; it can only be inferred from the outcome that results from his actions. This outcome depends also on a random state of nature. A bad outcome can be due either to improper actions by the agent or an unfortunate state of nature. Thus the agent can cover dysfunctional behavior behind the state of nature. As a result, optimal risk sharing cannot generally be attained because it does not provide appropriate incentives for action. The task then becomes to find a second-best solution, which optimally balances gains from risk spreading with gains from action incentives.

Examples of moral hazard are abundant and several papers have been written which address the issue. The classical example of moral hazard comes from insurance. If an agent is provided with perfect insurance he looses his incentives for taking preventive actions. This was observed by Arrow [1965]. An analysis of various special insurance models has since been undertaken by, among others, Kihlstrom and Pauly [1971], Pauly [1974], and Zeckhauser [1970]. The main conclusions in these papers exhibit the second-best nature of the problem.

Moral hazard in contracting for labor services has been analyzed

by Stiglitz [1974], [1975] and Noreen [1976]. In his 1974 paper, Stiglitz studies incentives and risk sharing in sharecropping, using a general equilibrium model of a competitive agricultural economy. Contracts are restricted to be linear. Many properties of the equilibrium contracts are derived, among them the following: if effort is unobservable, the equilibrium is inefficient compared to the standard of perfect information (except if workers are risk neutral); workers always receive a positive share of the output, which explains the sharecropping arrangement from the incentive point of view; there is no presumption in a general equilibrium model that sharecropping reduces effort from what it would have been under a wage system with enforceable contracts.

In a partial equilibrium framework Stiglitz [1975] compares piece vs. time rate payment systems in employment contracting. He exhibits how the optimal contract depends on attitudes towards risk, effort supply elasticities, the uncertainties involved and the supervision employed. Noreen's [1976] analysis is a partial extension of Stiglitz's. His major result shows that including options in addition to stock and fixed salary in the compensation package for executives, will result in Pareto improvements.

Another paper on restricted forms of employment contracts is by Keren [1968]. Keren studies the use of simple step functions as compensation schemes. These involve a specification of a target level for the outcome such that when the outcome exceeds the target, a bonus is paid. Keren asks what target will elicit maximal effort from the

worker. Thus only one point on the Pareto frontier is picked up. A characterization of the whole frontier appears substantially more complicated.

Spence and Zeckhauser [1971] and Ross [1973] have derived a characterization of a Pareto optimal general sharing rule under conditions of moral hazard. However, their analysis is incorrect at a rather fundamental point. Both assume that the optimal sharing rule is differentiable and proceed to characterize it using the calculus of variations. But Gjesdel [1976] has shown that the optimal solution may very well be nondifferentiable and, in fact, the first-best solution can occasionally be attained by using nondifferentiable sharing rules.

This observation has inspired our analysis. The formulation used by both Spence and Zeckhauser and Ross does not lend itself to an analysis of nondifferentiable sharing rules. In order to be able to study these, we have formulated the problem differently. This has also been done by Mirrlees [1974,]976], and some of our main conclusions coincide. We will, however, emphasize rigor and study in detail the existence of an optimal solution (for reasons that will become evident shortly), as well as the validity of the characterization of such a solution. Furthermore, we extend the analysis to situations where, in addition to the outcome, other signals about either the agent's action or the state of nature are observed. The main theorem provides a necessary and sufficient condition for such additional signals to be of value and included in the contract. This

condition requires that the signals provide information about the agent's action beyond what can be inferred from the outcome alone. It is a condition stated purely in terms of the relationship between the action and the probability distributions of the outcome and other observables. The result is a substantial extension of the analysis of monitoring provided by Harris and Raviv [1976].

4.2 Sharing a Jointly Produced Outcome

4.2.1 A Single Measure

There are n agents. Each agent i takes a nonobservable action $a_i \in A_i \subseteq R^1$, with a private (possibly nonmonetary) return $f_i: A_i \to R^1$. Together their actions result in a joint monetary outcome $x: A \to R^1$, which must be allocated among the agents. Here $A = \underset{i=1}{\overset{n}{\times}} A_i$, and we will write $a = (a_1, \ldots, a_n) \in A$. We assume an agent's preference function can be described as the sum of his private return $f_i(a_i)$ and his share in the outcome $x(a)$; i.e., it is additively separable and linear in money.[1]

The question is whether there is a way to share x so that the resulting noncooperative game between the agents has a Nash equilibrium which is Pareto optimal. That is, do there exist sharing rules $s_i(x)$, $i = 1, \ldots, n$, such that we have budget-balancing,

$$\sum_{i=1}^{n} s_i(x) = x, \qquad \forall x \in R^1, \tag{4.1}$$

and the noncooperative game with payoffs

$$(4.2) \qquad f_i(a_i) + s_i(x(a)) \quad \text{for} \quad i = 1, \ldots, n,$$

has a Nash equilibrium a*, which satisfies the condition for Pareto optimality,

$$(4.3) \qquad a^* = \underset{a \in A}{\operatorname{argmax}} \sum_{i=1}^{n} f_i(a_i) + x(a).$$

The answer is in the negative if we make the following assumptions:

A1. There exists a unique Pareto optimal solution $a^* = (a_1^*, \ldots, a_n^*)$, and a* belongs to the interior of A.[2]

A2. The functions x and f_i, i = 1, ..., n, are differentiable.

A3. $\frac{\partial x(a^*)}{\partial a_i} \neq 0$ for all i.

Condition A3 expresses that there is a genuine dependence between the agent's decisions at the optimum.

Theorem 4.1: Assume A1-A3. Then there do not exist sharing rules $s_i(x)$, which satisfy (4.1) and for which the Pareto optimal decision a* is a Nash equilibrium in the noncooperative game with payoffs (4.2).

Proof: Let s_i, i = 1, ..., n, be arbitrary sharing rules which staisfy (4.1). We will show that the assumption that a* is a Nash equilibrium will lead to a contradiction.

From the definition of a Nash equilibrium we have:

$$(4.4) \qquad f_i(a_i) + s_i(x(a_i, a^{*i})) \leq f_i(a_i^*) + s_i(x(a^*)) \quad \forall a_i \in A_i.$$

Let $\{x^\ell\}$ be a strictly increasing sequence of real numbers, which converges to $x(a^*)$. Let $\{a_i^\ell\}$ be the corresponding n sequences, which satisfy

$$(4.5) \qquad x^\ell = x(a_i^\ell, a^{*i}) \qquad \forall i, \forall \ell.$$

Such sequences can be defined by assumption A3 (starting from a sufficiently large ℓ if necessary). Pareto optimality and A1 imply

$$f_i'(a_i^*) = -\frac{\partial x(a^*)}{\partial a_i} \quad \forall i.$$

This in turn implies, using (4.5),

$$f_i(a_i^\ell) - f_i(a_i^*) = x(a^*) - x(a_i^\ell, a^{*i}) + o(a_i^\ell - a_i^*) \quad \forall i, \forall \ell.$$

where $o(h)/h \to 0$ as $h \to 0$. Substitution into (4.4) gives

$$x(a^*) - x^\ell + o(a_i^\ell - a_i^*) \leq s_i(x(a^*)) - s_i(x^\ell) \qquad \forall i, \forall \ell.$$

Sum over i and use (4.1) to get

$$\sum_{i=1}^{n} \{x(a^*) - x^{\ell} + o(a_i^{\ell} - a_i^*)\} \leq 0 \qquad \forall \ell.$$

By the differentiability of x this can be written

$$(4.6) \qquad \sum_{i=1}^{n} \{- \frac{\partial x(a^*)}{\partial a_i} (a_i^{\ell} - a_i^*) + o(a_i^{\ell} - a_i^*)\} \leq 0 \qquad \forall \ell.$$

Since $x^{\ell} < x(a^*)$, by our choice of x^{ℓ}, the first term in the bracket is strictly positive in view of A3. This term dominates for large ℓ, which contradicts (4.6). Hence, the assumption that a^* is a Nash equilibrium has led to a contradiction and must be false.

Q.E.D.

The idea behind the proof is quite simple. If a^* were to be a Nash equilibrium, each agent should at the margin carry the total social loss from a decrease in x. But this is not possible, since the shares have to sum up to x only. Hence, budget-balancing plays again a crucial role in destroying the efficiency of a Nash equilibrium. But there is another reason in conjunction with budget-balancing, namely that we cannot discern the actions taken by the agents from the single outcome measure x. Because an agent can cover an improper action behind the uncertainty of who was at fault, and because all agents cannot be penalized sufficiently for a deviation in the outcome, each agent always has an incentive to capitalize on this control deficiency.

The problem described above is likely to arise in many real

world situations. Examples include labor-managed enterprises, farm cooperatives, management teams, and professional-services firms like CPA partnerships. In all cases labor and ownership are integrated, and this results in insufficient supply of productive inputs like effort.

This is the starting point for Alchian and Demsetz's [1972] reasoning. They argue that striving for higher efficiency will result in an organizational change. A monitor will be hired to measure the marginal productivity of each agent, and to the extent he is successful, workers will get paid their marginal product and efficiency is restored. This requires that the monitor is equipped with the right to terminate memberships in the team in order to induce proper behavior. But what guarantees that the monitor will provide the right amount of effort for monitoring? Alchian and Demsetz's solution is to give the monitor the title to net earnings, net of payments to the other agents. In this way the monitor becomes effectively the owner of the firm.

We have omitted details, but this is the main line of reasoning behind Alchian and Demsetz's theory of the classical capitalist firm. Notice though that adding the monitor to the team will transform it into a new augmented team. Why do we not have the same problems with inefficiency again? The reason is that there are now several measures of the outcome available due to monitoring. But only if monitoring will discern perfectly the agents' deviations from proper actions, will efficiency be achieved. We will return to the question

of how rich the measurement system should be in order to guarantee efficiency. Let us, however, first discuss the other alternative solution to the problem: elimination of the budget-balancing condition.

If budget-balancing is not a constraint, one can make the efficient outcome a Nash equilibrium by giving each agent the total share of the outcome. This is the only smooth scheme which will work, as is easily seen from the two conditions:

$$\text{PO}: f_i'(a_i^*) + \frac{\partial x(a^*)}{\partial a_i} = 0, \qquad \forall i,$$

$$\text{NE}: f_i'(a_i^*) + s_i'(x(a^*)) \cdot \frac{\partial x(a^*)}{\partial a_i} = 0, \qquad \forall i.$$

These conditions imply $s_i'(x(a^*)) = 1$, and consequently $s_i(x) = x + k_i$, $\forall i$ at $x(a^*)$, where k_i is an arbitrary constant.

From a pragmatic point of view such linear sharing rules may not be desirable. If the agents for some reason choose an action vector $a \neq a^*$ such that $x(a) > x(a^*)$, there will be insufficient funds to compensate them and the scheme becomes infeasible. A more appropriate scheme would be discontinuous at $x(a^*)$. For instance, letting $s_i(x)$ be a step function, which pays the efficient share for $x \geq x(a^*)$ with a drop in returns (sufficient to induce a^* as a Nash equilibrium) if $x(a^*)$ is not achieved, will work. Such group incentives, where all agents are penalized since the ones at fault cannot be discerned, are found for instance in contracting with labor teams. Usually it takes

the form of a flat wage for team members, with a bonus paid if the efficient output level is attained; (whether we view the discontinuity as a bonus or a penalty is, of course, a matter of taste).

We find then that no monitoring is needed in order to achieve efficiency, when we do not have state uncertainty. All it takes is to relax budget-balancing. The most natural way of achieving this is to separate ownership and labor, that is, introduce an owner to the organization who does not provide any productive inputs, but merely picks up the residual of the nonbudget-balancing sharing rule, which induces efficient actions. This explains the emergence of capitalist firms somewhat differently than Alchian and Demsetz. The fact that stock owners do not generally exercise very close monitoring of the behavior of managers, only of managers' performance measured by the total outcome, supports our theory.

This is not to say that monitoring can generally be abandoned. Its role is important in two ways. First, use of discontinuous bonus (or penalty) schemes leads to an infinite set of Nash equilibria. All will result in an efficient production level, but only one will lead to efficient distribution of labor supply. If one agent undersupplies labor, this will be fully compensated by others up to a certain limit. For this reason monitoring may play a role as a disciplinary tool for the team even without state uncertainty.

Secondly, and more importantly, the presence of uncertainty may change the picture drastically. Discontinuous contracts can become quite impractical and inefficient if the outcome will vary beyond the

agents' control. We will see in Section 4.3 that monitoring becomes an essential ingredient in the organization under such circumstances.

At this point a comment on Kleindorfer and Sertel [1976a] is appropriate. They study team production using the same model we have presented, but restrict themselves to linear sharing rules such that the sum of the shares is less than or equal to one. The owner, who provides no inputs himself, optimizes over this set of linear shares, and selects that Nash equilibrium which gives him the highest residual. The result is inefficient in the sense described by Theorem 4.1. As we have argued above, efficiency can be achieved only by going outside the class of linear schemes. Hence, it seems that the restriction to linear schemes is not very well motivated in this case. On the other hand, Kleindorfer and Sertel [1976b] apply the same model (without an owner) to a labor-managed firm or cooperative. Then the use of linear schemes is no restriction, since it can be shown that any Nash equilibrium which can be attained with general sharing rules that balance the budget, can likewise be attained with constant shares summing up to one.

4.2.2 Additional Measures

Our conclusion from the previous section is that it is insufficient to observe the total outcome if one wants to achieve an efficient noncooperative equilibrium when budget-balanding is imposed. If budget-balancing cannot be relaxed the alternative is to observe additional signals about the agents' actions or, as we will view it, get a more detailed account of the outcome measure x.

Let x consist of the sum of m measures $x_k: A \to \mathbb{R}^1$, $k = 1, \ldots, m$, i.e.,

$$(4.7) \qquad \sum_{k=1}^{m} x_k(a) = x(a), \qquad \forall a \in K.$$

We call the set of functions x_k, an accounting system. Based on this accounting system, we can design an allocation mechanism, which is a set of sharing rules $s_i: \mathbb{R}^m \to \mathbb{R}$, $i = 1, \ldots, n$, satisfying:

$$(4.8) \qquad \sum_{i=1}^{n} s_i(x_1, \ldots, x_m) = x \qquad \forall x.$$

The pair consisting of an allocation mechanism and an accounting system will be called a control system. If a control system leads to a Nash equilibrium at the Pareto optimal action a*, we say that the control system is acceptable. If an accounting system is rich enough so that an acceptable control system can be built upon it, we say that the accounting system is sufficient.

With this terminology our problem can be posed as follows: find the conditions under which an accounting system is sufficient. The result from the previous section (Theorem 4.1) was that the total outcome alone is an insufficient accounting system.

The reason why a richer set of measures may help to control the agents better, is, of course, that several measures generally make it possible to infer more about individual actions. In the limit, a sufficiently rich accounting system may reveal exactly the actions of

the agents. In that case an acceptable control system can easily be constructed. We state this formally in the following:

Theorem 4.2: If the accounting system is a one-to-one mapping from decisions to outcomes, then it is sufficient.

Proof: The measures will reveal each agent's decision and so we can make the sharing rules directly dependent on these decisions. Let a^* be a Pareto optimal decision. Let the sharing rules at a^* be $s_i(a^*)$, $i = 1, \ldots, n$. We will show that for any $a \in A$, the outcome can be shared so that:

$$f_i(a_i) + s_i(a) \leq f_i(a_i^*) + s_i(a^*) \quad \forall i. \tag{4.9}$$

This clearly implies our claim.

Let $a \in A$ be arbitrary, and suppose (4.9) cannot be achieved. That implies there exist sharing rules $\hat{s}_i(a)$ such that

$$f_i(a_i) + \hat{s}_i(a) \geq f_i(a_i^*) + s_i(a^*), \quad \forall i, \tag{4.10}$$

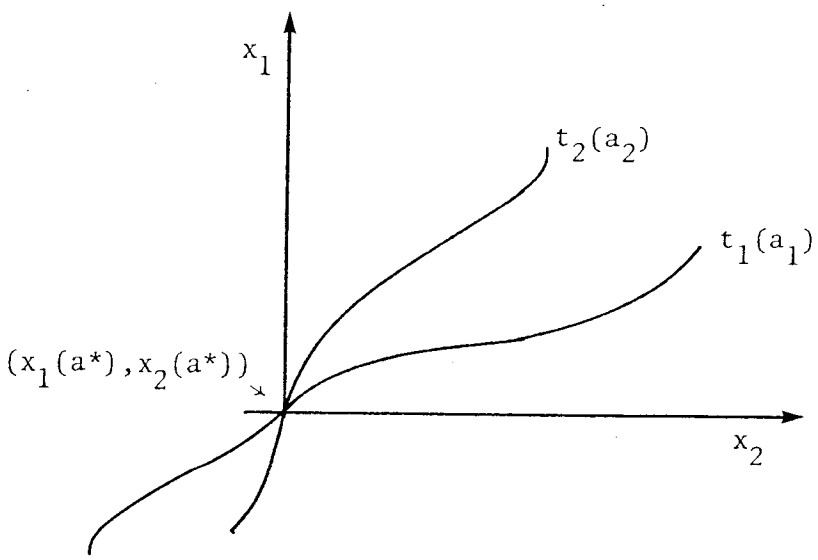
with strict inequality for at least one. Add (4.10) over all i's to get

$$\sum_{i=1}^{n} f_i(a_i) + x(a) > \sum_{i=1}^{n} f_i(a_i^*) + x(a^*),$$

using (4.1). This contradicts the Pareto optimality of a^*. Q.E.D.

All Theorem 4.2 says is that if actions are observable or possible to infer with certainty, one can achieve efficiency. The payoffs of the noncooperative game can be redistributed in such a manner that the most desirable outcome is the only Nash equilibrium.

The assumption of observability is quite strong in Theorem 4.2 and can be weakened. It suffices that we can detect when an agent is the only one who deviates from the optimum. This will be possible if and only if the curves $t_i(a_i) = (x_1(a_i, a^{*i}), \ldots, x_m(a_i, a^{*i})) \in \mathbb{R}^m$, $i = 1, \ldots, n$, differ as illustrated in the figure below.



We say that an accounting system is <u>independent at a*</u> if and only if there does not exist an $a \in A$, $a \neq a^*$ such that $t_1(a_1) = \ldots = t_n(a_n)$.

Theorem 4.3: An independent accounting system is sufficient.

Proof: Let $s_i(x_1(a^*), \ldots, x_n(a^*))$, $i = 1, \ldots, n$, be an arbitrary split of $x(a^*)$, which satisfies (4.1). Define sharing rules s_i along the t_i-curves as follows:

$$(4.11) \qquad s_i(t_i(a_i)) = s_i(x_1(a^*), \ldots, x_m(a^*)) + x(a_i, a^{*i}) - x(a^*)$$

for $i = 1, \ldots, n$, and the others arbitrary but so that (4.1) holds. This is possible by our assumption of independence. With such a choice the agent's objective coincides with the social objective when others stick to their efficient action a^{*i}. Hence, the agent's best response against a^{*i} will be a_i^* by definition of a^*.

Q.E.D.

We notice that one measure does not constitute an independent accounting system. Independence is also a necessary condition in the sense that a sufficient accounting system has to be independent at least in the neighborhood of a^*. We further notice that if actions of the agents are perfect substitutes of each other, then no accounting system can be independent, since by definition of substitutability, for any a_i, there exists an a_j for each $j \neq i$, such that $t_i(a_i) = t_j(a_j)$.

From Theorem 4.3 we see that two measures may well be sufficiently rich to reveal individual deviation. However, if we make

the assumption that x_k's are monotone in actions, e.g.,

$$\text{A4} \qquad \frac{\partial x_k(a)}{\partial a_i} \geq 0 \qquad \text{for every } a \in A,\ i = 1, \ldots, n,$$

and constrain ourselves to differentiable and monotone sharing rules,

$$(4.12) \qquad \frac{\partial s_i}{\partial x_k}(x_1, \ldots, x_n) \geq 0 \qquad \forall x, \forall i,$$

then at least n measures are needed.

Theorem 4.4: Assume A1-A4. Then the accounting system has to include at least n measures if one wants to construct a monotone acceptable control system.

Proof: By (4.7) and (4.8) we have

$$(4.13) \qquad \sum_{i=1}^{n} s_{ik} = 1$$

where $s_{ik} = \frac{\partial s_i}{\partial x_k}(x_1(a^*), \ldots, x_n(a^*))$.

By (4.3) and (4.7)

$$(4.14) \qquad f_i' + \sum_{k=1}^{m} x_{ki} = 0 \qquad \forall i,$$

where f_i' is evaluated at a^*, and $x_{ki} = \frac{\partial x_k}{\partial a_i}(a^*)$. From the Nash equilibrium property of a^*

$$f_i' + \sum_{k=1}^{m} s_{ik} \cdot x_{ki} = 0. \tag{4.15}$$

Combining (4.14) and (4.15) yields

$$\sum_{k=1}^{m} x_{ki}(1 - s_{ik}) = 0. \tag{4.16}$$

By (4.12), (4.13) and A4,

$$x_{ki}(1 - s_{ik}) = 0 \quad \forall i, \forall k.$$

By A3, $x_{ki} > 0$ for at least one k for a given i, say k_i. Then $s_{ik_i} = 1$, which implies $s_{jk_i} = 0$, $\forall_j \neq i$, by (4.12) and (4.13). Hence, there must be at least n measures, since each agent is given the full share in at least one. Q.E.D.

The assumption of monotonicity is rather natural to make if we think of the x_k's as monetary outcomes which improve with, say, increased effort. In practice most sharing rules are monotone. All agents get a positive share in the outcome. Under such circumstances efficiency can be achieved only if each agent is in charge of his own account. Moreover, the proof shows that it must be that his action does not affect the other agents' accounts. In other words, only when the whole system can be decoupled and externalities removed can we achieve efficiency (compare to Section 3.4).

The conclusion is that if budget-balancing is required, the only way to reduce inefficiencies is to create a richer accounting system which better discerns individual deviations. Two measures may be sufficient, but if they are monotone and we want a monotone allocation mechanism, then n independent measures are needed which in effect decouple the organization. The desire to decouple the organization is familiar from responsibility accounting. The analysis supports the widely accepted accounting principle that managers should be able to control the measures that are used for evaluation of their performance (see Horngren [1972], Chapter 6 on responsibility accounting and motivation).

We have not discussed the possibility that some decision, say an allocation of the firm's resources, may make agents' actions dependent. If one tries to promote goal congruence, in order to guarantee an efficient allocation of resources by giving each agent a share in the firm's outcome, this will again lead to insufficient supply of effort. It is interesting to note that Groves' scheme is able to get around this problem. By effectively decoupling the organization, it can assure both otpimal allocation of resources and efficient supply of effort. Once the allocation of resources is determined, each agent is in charge of his own account as required for efficiency (see Section 3.2.2).

4.3 Principal-Agent Relationship under Uncertainty

4.3.1 Two Problem Formulations

A principal and an agent have to share a random outcome $x(a,z)$, which depends on the nonobservable action $\underline{a}$ of the agent, and on the uncertain state of nature $\tilde{z}$. We assume that they have homogeneous beliefs about $\tilde{z}$, embodied formally in a probability space (Z, F, P), and focus on the moral hazard aspect of the problem that arises when the action is not observed by the principal. In particular, we are interested in the characteristics of the agent's share $s(x)$ under (constrained) Pareto optimality.

The principal's utility is over wealth alone, $G(w)$; the agent's utility is over wealth and actions, $U(w,a)$. We will assume:

B1. $a \in A$, a compact subset of $\mathbb{R}$.

B2. $G: \mathbb{R} \to \mathbb{R}$, $U: \mathbb{R}^2 \to \mathbb{R}$ are twice continuous differentiable;

$$G' > 0,\ G'' \leq 0,\ U_w > 0,\ U_{ww} < 0,\ U_a \leq 0,\ U_{aa} \leq 0.$$

Mostly we will be working with a situation in which the action is a productive input of the agent, most conveniently thought of as effort. In that case it is natural to assume that $U_a < 0$, and $x_a(a,z) \geq 0$ for every $z \in Z$, but we will not yet make these assumptions explicit in order to cover a model by Ross [1973], in which $U_a \equiv 0$.

Constrained Pareto-optimal action-sharing rule pairs (a,s) can be generated by solving:[3]

$$(4.17) \qquad \max_{a,s(x)} \int\{G(x(a,z) - s(x(a,z))) + \lambda \cdot U(s(x(a,z)), a)]dP(z),$$

$$(4.18) \qquad \text{s.t.} \quad a \in \operatorname*{argmax}_{a\in A} \int U(s(x(a,z)),a)dP(z).$$

Assuming that s is differentiable and the agent's maximizing action in (4.18) is unique and interior in A, we can replace (4.18) by the first-order condition:

$$(5.19) \qquad \int \{U' \cdot s' \cdot x_a + U_a\}dP(z) = 0.$$

To characterize the optimal sharing rule s, one fixes a and solves for s using a standard calculus of variations argument. This is the approach taken in Ross [1973] and Spence and Zeckhauser [1971].

Two main assumptions have to be made to validate the procedure above. The first is that an optimal solution exists, and the second one is that this solution is differentiable. As has been shown in Gjesdal [1976] and Mirrlees [1974], both assumptions may quite generally be false. Gjesdal studied cases in which the distribution of z has compact support and $x_a(a,z) > 0$ for all z. To illustrate his ideas we look at the following simple example:

Example 4.1

$$x(a,z) = k \cdot a + z, \quad z \sim \text{unif}(-1,1), \quad a \geq 0, \quad k > 0.$$

$$U(w,a) = U(w) - V(a) = \log(w) - \frac{1}{2} \cdot a^2,$$

$$G(w) = w.$$

A first-best solution $(\bar{a},\bar{s})$ satisfies:[4]

$$\bar{a} = k/\lambda, \quad \bar{s}(x) = \lambda, \quad \text{for all} \quad x;\lambda > 0.$$

Define the following sharing rule:

$$s(x) = \lambda, \quad \text{if} \quad x \geq g \equiv \bar{a} - 1,$$
$$= v, \quad \text{if} \quad x < g; \; v < \lambda .$$

With s as a payment schedule, the agent chooses a so as to maximize:

$$U(\lambda) \; (1 - F(g,a)) + U(v) \cdot F(g,a) - V(a)$$

where F(g,a) is the probability that x is below g, when the agent takes action a; i.e.,

$$F(g,a) = \begin{cases} 0, & \text{if} \quad a \geq \bar{a}, \\ \dfrac{\bar{a} - a}{2}, & \text{if} \quad a \in [\bar{a} - 2, \bar{a}], \\ 1, & \text{if} \quad a \leq \bar{a} - 2. \end{cases}$$

Now, if v is chosen such that:

$$[\log \lambda - \log v] > 2 \cdot \bar{a} = 2k/\lambda,$$

then the agent's optimal act under s equals $\bar{a}$. Moreover, he will be certain to receive λ, when he takes this action. Hence, the first-best solution can be attained by the nondifferentiable sharing rule s. It is easily seen that no differentiable rule can precisely attain the first-best solution, and so, at least for this example, the assumption of differentiability is too restrictive. □

It is quite clear what is driving the result in the example. If the agent takes an act below $\bar{a}$, then there is a positive probability that he will be detected, and by penalizing detection sufficiently the agent will not take this risk. On the other hand, and this is crucial for actually attaining the first-best solution, the agent can avoid any penalties by taking the correct action.

We have added the parameter k in the example to illustrate that if x is not very sensitive to a change in action, then it may take very

high penalties to actually induce the agent to take the correct action $\bar{a}$ and achieve a first-best solution.

At first sight it may seem that the compact support of z is a mathematical trick which makes the example work. Partly that is true. A first-best solution will never be attainable if the support is the real line, but as we will see, arbitrarily close approximations may be possible quite generally. In any case, the example shows that differentiability is not to be taken for granted. In order to study non-differentiable sharing rules, the problem has to be formulated differently.

An Alternative Formulation

Rather than viewing x as a function of $\underline{a}$ and z explicitly as in (4.17)-(4.19), we can look directly at the distribution of x as a function of $\underline{a}$. This distribution is denoted F(x,a). The relationship between F(x,a) and the distribution of z when we employ some common production functions, has been recorded in Appendix 4A for further use.

We will assume:

B3. F(x,a) has a density function f(x,a), which is twice continuously differentiable in $\underline{a}$.

Later we will relax this condition so as to allow for mass points in the distribution of x.

B4 (i) $\int |f_a(x,a)|dx < \infty$, $\forall a \in A$,

(ii) $\int |f_{aa}(x,a)|dx < \infty$, $\forall a \in A$,

(iii) $\int |G(x) \cdot f(x,a)|dx < \infty$,

(iv) $\int |G(x) \cdot f_a(x,a)|dx < \infty$.

B4 guarantees (by bounded convergence) that we can differentiate under the integral sign as needed in the sequel. Our problem can now be formulated as follows:

$$(4.20) \qquad \max_{a,s(x)} \int \{G(x - s(x)) + \lambda \cdot U(s(x),a)\} f(x,a)dx,$$

$$(4.21) \qquad \text{s.t.} \quad a \in \operatorname*{argmax}_{a \in A} \int U(s(x),a) f(x;a)ds$$

Since we will restrict ourselves to bounded sharing rules $s(x)$, the problem is well-defined by B4. Furthermore, (4.21) can be replaced by:

$$(4.22) \qquad \int \{U(s(x),a) f_a(x,a) + U_a(s(x),a) f(x,a)\}dx = 0.$$

The point is that, in order to be able to write out (4.22), we need not assume that $s(x)$ is differentiable. It suffices that it is bounded and measurable (since U and U_a are continuous) when we have assumed B2-B4. This is one of the major advantages of the formulation (4.20)-(4.22) compared to (4.17)-(4.19), but there are others as well as we will see shortly.[5]

4.3.2 Existence of an Optimal Solution

We now turn to the issue of existence of an optimal solution. We will prove that (4.20)-(4.21) has an optimal solution (a^*,s^*) when we restrict ourselves to sharing rules that belong to one of the following two classes of functions:

$$S_1 = \{s:\mathbb{R} \to [c,d] \mid s \text{ nondecreasing}\},$$

or

$$S_2 = \{s:\mathbb{R} \to [c,d] \mid s \text{ has modulus of continuity } \delta(\varepsilon)\}.$$

Here $\delta:\mathbb{R}_+ \to \mathbb{R}_+$ is a fixed function with $\lim_{\varepsilon \to 0} \delta(\varepsilon) = 0$, and $s \in S_2$ if $|x-y| < \delta(\varepsilon)$ implies $|s(x) - s(y)| < \varepsilon$. S_2 is called an equi-continuous family of functions. An example of an equi-continuous family of functions is the set of all functions which satisfy a Lipschitz condition of the form $|s(x) - s(y)| \leq M \cdot |x-y|$. In this case $\delta(\varepsilon) = \varepsilon/M$.

The reason we restrict ourselves either to S_1 or S_2 is that these classes can be shown to be sequentially compact as is done in Appendix 4B. Notice that in both classes s is bounded. This boundedness is instrumental for existence and we will see why when we get to Section 4.3.5.

For ease of notation we define the following functionals:

$$E^P(a,s) = \int G(x - s(x))f(x,a)dx,$$

$$E^A(a,s) = \int U(s(x),a)f(x,a)ds,$$

$$B(a,s) = E^P(a,s) + \lambda \cdot E^A(a,s).$$

These are mappings from $\mathbb{R}xS \to \mathbb{R}$, where S is the subspace of functions in which we are optimizing. For a particular choice of s, the agent is assumed to maximize his expected utility $E^A(a,s)$. By boundedness of s and B2-B4, the partial derivatives $E^A_a(a,s)$ and $E^A_{aa}(a,s)$ exist and are continuous. Since A is compact, there will exist a solution to the agent's problem. Define the solution correspondence

$$a(s) = \{a\varepsilon A \mid a\varepsilon \underset{a\varepsilon A}{\text{argmax}}\ E^A(a,s)\}.$$

We notice that a(s) is a closed set by the continuity of $E^A(a,s)$, and compact because it is a subset of A. If a(s) contains more than one point we have to make a further assumption about the agent's behavior. We will assume:

B5. If a(s) is multi-valued the agent chooses an $a\varepsilon a(s)$ which maximizes the principal's objective function $E^P(a,s)$.

From the discussion above such maximizing elements in a(s) exist as long as $E^P(a,s)$ is continuous in a, which is guaranteed by B4-(iii) and the fact that s(x) is bounded. Finally, there may be more than one $a\varepsilon a(s)$ which maximizes $E^P(a,s)$. To break such ties an arbitrary selection is made. This will define the agent's response function:

$$a_{max}: S \to A.$$

With this notation, (4.20)-(4.22) can be rewritten as:

$$(4.23) \qquad \max_{s\in S, a\in A} \; B(a,s) \equiv E^P(a,s) + \lambda \cdot E^A(a,s),$$

$$(4.24) \qquad \text{s.t.} \quad E^A_a(a,s) = 0;$$

or alternatively:

$$(4.25) \qquad \max_{s\in S} J(s) \equiv B(a_{max}(s),s).$$

In the appendix it is shown that J(s) is an u.s.c. function (in an appropriate topology) and the proof of the main existence theorem follows by standard arguments:

Theorem 4.5: Let $S = S_1$ or S_2. Assume B1-B5. Then there exists an optimal solution to Problem (4.25), (a^*,s^*), with $a^* = a_{max}(s^*)$.

Proof: See Appendix 4B.

Theorem 4.5 says that if we restrict ourselves either to non-decreasing and bounded sharing rules or to a family of bounded equi-continuous functions, then there exists a Pareto optimal solution.

4.3.3 A Characterization of s*

Define the following class of functions:

$$S_3 = \{s:\mathbb{R} \to [c,d] \mid s \text{ is measurable}\},$$

and assume:

B6. There exists an optimal solution, (a^*,s^*), to problem (4.25) when $S = S_3$, such that a^* is in the interior of A and uniquely maximizes $E^A(a,s^*)$.

S_3 is not sequentially compact in an appropriate topology, and for this reason B6 is necessary. But given the nature of the examples of nonexistence that we will discuss later, we believe that B6 is not very restrictive. We will first characterize an optimal solution $s^* \varepsilon S_3$ and then discuss implications of further restrictions to either S_1 or S_2 (which are subsets of S_3).

A necessary condition for an optimal sharing rule can be derived using a first-order approximation of J(s) in the neighborhood of s*. Such an approximation is given in Luenberger [1968] (proposition 9.6.1). Since our assumptions do not exactly match Luenberger's, we will reproduce his proof to show that they are sufficient for our purposes. Let (a^*,s^*) be a solution satisfying B6, $s \varepsilon S_3$ arbitrary, and $h \varepsilon \mathbb{R}$. Define:

(4.26) $$s_h(x) = h \cdot s(x) - (1-h)s^*(x), \quad \text{for every } x \varepsilon \mathbb{R},$$

Also, define the Lagrangian:

$$L(a,s,\mu) = B(a,s) + \mu \cdot E^A_a(a,s).$$

Then we have the following approximation lemma:

Lemma 4.6: Assume B2-B6. Let $\mu^* \in \mathbb{R}$ satisfy:

$$B_a(a^*,s^*) + \mu^* \cdot E^A_{aa}(a^*,s^*) = 0. \tag{4.27}$$

Then,

$$J(s^*) - J(s_h) = L(a^*,s^*,\mu^*) - L(a^*,s_h,\mu^*) + o(h), \tag{4.28}$$

in a neighborhood of $h = 0$. Here $a^* = a(s^*)$, and $o(h)/h \to 0$ as $h \to 0$.

Proof: In view of B6 it is clear that there is an interval $I = (-\delta,\delta)$ s.t. the correspondence $a(s_h)$ is single-valued for $h \in I$. Since, $E^A_a(a,s_h)$, as a function of $\underline{a}$ and h, is differentiable w.r.t. both arguments, we have for some $K > 0$,

$$|a(s_h) - a(s^*)| \leq K \cdot |h|, \quad \text{for} \quad h \in I. \tag{4.29}$$

We can write:

$$J(s^*) - J(s_h) = B(a^*,s^*) - B(a(s_h),s_h)$$

$$= B(a^*,s^*) - B(a^*,s_h) + B(a^*,s_h) - B(a(s_h),s_h)$$

$$= B(a^*,s^*) - B(a^*,s_h) + B_a(a^*,s^*)(a^* - a(s_h))$$

$$+ [B_a(a(s_h),s_h) - B_a(a^*,s^*)](a^* - a(s_h)) + o(h)$$

$$= B(a^*,s^*) - B(a^*,s_h) + B_a(a^*,s^*)(a^* - a(s_h)) + o(h).$$

For the last two steps we have used the fact that B_a is continuous (which follows from B4 and the boundedness of s^* and s_h), and (4.29). We also have by continuity of E^A_{aa} and (4.29):

$$E^A_{aa}(a^*,s^*)(a^* - a(s_h)) = E^A_{aa}(a^*,s_h)(a^* - a(s_h)) + o(h)$$

$$= E^A_a(a^*,s_h) - E^A_a(a(s_h),s_h) + o(h)$$

$$= E^A_a(a^*,s_h) - E^A_a(a^*,s^*) + o(h) \quad \text{(by (4.24))}.$$

The result follows by using (4.27).

Q.E.D.

(4.28) says that a necessary condition for optimality is a stationary Lagrangian. This result could, of course, have been obtained by maximizing $B(a^*,s_h)$ subject to $E^A_a(a^*,s_h) = 0$, but for later use we prefer

the derivation above, as it explicitly shows how the change in J(s) can be approximated once we have a μ that satisfies (4.27). We can use the lemma to prove the first characterization theorem:

Theorem 4.7: Assume B2-B6. Let (a^*,s^*) be an optimal solution satisfying B6. Then $s^*(x)$ satisfies one of the three conditions below for almost every $x\varepsilon\mathbb{R}$:

$$\text{(i)} \quad H_t(a^*,s^*(x),x,\mu^*) = 0, \quad s^*(x)\varepsilon(c,d),$$

$$\text{(ii)} \quad H_t(a^*,c,x,\mu^*) \leq 0, \quad s^*(x) = c,$$

$$\text{(iii)} \quad H_t(a^*,d,x,\mu^*) \geq 0, \quad s^*(x) = d,$$

where H_t is the partial derivative of the Hamiltonian:

$$H(a,t,x,\mu) = G(x-t) \cdot f(x,a) + \lambda \cdot U(t,a) \cdot f(x,a)$$

$$+ \mu \cdot U(t,a) \cdot f_a(x,a) + \mu \cdot U_a(t,a) \cdot f(x,a),$$

and μ^* satisfies (4.27).

Proof: Suppose the claim were false. Then there would exist a set of positive measure such that one of the three conditions would be false. Let us assume it is (ii). Then $H_t > 0$ and $s_t(x) = c$ on a set of positive measure. This implies (by continuity of the Lebesque-measure) that there is a set X of positive measure m(X), such that:

$$H_t(a^*,c,x,\mu^*) > \varepsilon, \quad \text{for} \quad x\varepsilon X.$$

Define $s(x) = s^*(x) + \delta$, $\delta > 0$, and

$$s_h(x) = h \cdot s(x) + (1-h)s^*(x), \quad \text{for} \quad x\varepsilon X,$$

$$= s^*(x) \quad \text{otherwise.}$$

The Lagrangian $L(a^*,s_h,\mu^*)$ will then be differentiable w.r.t. h and we get for $h > 0$:

$$L(a^*,s_h,\mu^*) - L(a^*,s^*,\mu^*) =$$

$$\int_X H_t(a^*,s^*(x),x,\mu^*)(s(x) - s^*(x))dx \cdot h$$

$$+ o(h) \geq m(X) \cdot \varepsilon \cdot \delta \cdot h + o(h).$$

Taking h small enough yields $J(s_h) - J(s^*) > 0$, by (4.28) and since $s_h \varepsilon S_3$, we have a contradiction to the optimality of s^*. Hence (ii) cannot be biolated on a set of positive measure. A similar argument shows that (iii) cannot be violated on a set of positive measure. Finally, the argument for (i) is either identical to case (ii) or (iii).

Q.E.D.

Remark: Along the same lines one could have proved that the Hamiltonian has to be point-wise maximized (not just point-wise stationary), but the Lipschnitz-condition (4.29) would have required some

additional assumptions, since we would have had to work in a function space with L_1-norm (see Luenberger [1968]).

From now on we will restrict ourselves to the case where the agent's utility function is separable and we write:

$$U(w,a) = U(w) - V(a).$$

In that case we get the following characterization:

Corollary 4.8: Assume B2-B6. Let (a^*,s^*) be an optimal solution satisfying B6, and let the agent's utility function be separable. Then $s^*(x)$ will maximize the Hamiltonian pointwise almost everywhere and:

$$\frac{G'(x - s^*(x))}{U'(s^*(x))} = \lambda + \mu^* \cdot \frac{f_a(x,a^*)}{f(x,a^*)} \tag{4.30}$$

if the equation has a solution $c \leq s^*(x) \leq d$. Otherwise:

$$\text{(i)} \quad s^*(x) = c, \quad \text{if} \quad G'(x - c) - (\lambda + \mu^* \cdot \frac{f_a}{f})U'(c) > 0,$$

$$\text{(ii)} \quad s^*(x) = d, \quad \text{if} \quad G'(x - d) - (\lambda + \mu^* \cdot \frac{f_a}{f})U'(d) < 0.$$

Proof: We notice that with a separable utility function for the agent, the Hamiltonian is either concave or nonincreasing, which together with the previous theorem gives the claim.

Q.E.D.

(4.30) is our main characterization result (also found in Mirrlees [1976], without boundary conditions). It is easily interpreted in the light of optimal risk-sharing without moral hazard constraints (see Wilson [1968]). If $\mu = 0$, then $s^*(x)$ would correspond to optimal risk-sharing. As we will see in the next section this never happens when $V'(a) > 0$. Instead $\mu \neq 0$, and s^* will deviate from otpimal risk-sharing in order to induce proper incentives for action.

Suppose a is effort. Then $\mu > 0$ normally (which is equivalent to saying that the principal would like a higher effort level at the optimum). From (4.30), $\mu > 0$ implies that $s^*(x)$ lies above optimal risk-sharing when $f_a(x,a^*) > 0$ and below when $f_a(x,a^*) < 0$. This corresponds exactly with our intuition, since a raise in the share when $f_a > 0$ or a cut when $f_a < 0$, will induce the agent to supply more effort than he would under optimal risk-sharing.

A Comment on Ross [1973]

When $\mu \neq 0$, the optimal sharing rule is crucially dependent on the distribution of the state of nature, whereas the distribution plays no role in optimal risk-sharing (with homogeneous beliefs as we have here). The reason is, of course, that one wants to capitalize on the informational content of the outcome as a signal about the agent's action. For the reader who is familiar with Ross [1973], this stands in some contrast to his claim that one can assume without loss of generality that $\tilde{z}$ has a uniform distribution on (0,1). Technically, this statement is correct in the sense that the problem (4.17)-(4.19)

(which is the formulation Ross uses) can always be reduced to one where $\tilde{z}$ has a uniform distribution, by a proper change of variables. However, it is somewhat misleading to state (as Ross does) that the characterization of $s^*(x)$ is:

$$(4.31) \qquad \frac{G'(x - s^*(x))}{U'(s^*(x))} = \lambda + \mu \cdot \frac{d}{dz}\left(\frac{x_a}{x_z}\right),$$

which does not seem to depend on the distribution of $\tilde{z}$. The dependence is in this case hidden in the term $\frac{d}{dz}\left(\frac{x_a}{x_z}\right)$, since in (4.31) $x(a,z)$ does not stand for the original outcome function, but the transformed one which results after the change of variables.

It is important to recognize this fact both for a proper understanding of (4.31) and of Ross's further results. It is well-known (see Wilson [1969]) that $\mu = 0$ in (4.31) (and hence $s^*(x)$ will provide efficient risk-sharing) for all outcome functions $x(a,z)$ and all distributions of $\tilde{z}$, only if u and g belong to the class of utility functions with linear absolute risk-aversion. Since this class is quite restricted, Ross asks what outcome functions will yield efficiency regardless of the pair of utility functions (U,G). This class he derives by requiring

$$\frac{d}{dz}\left(\frac{x_a}{x_z}\right) = b(a),$$

from (4.31), and solving the partial differential equation. But notice that this equation is only relevant for a uniformly distributed $\tilde{z}$. So

the solution $x(a,z) = h(z \cdot b(a) - c(a))$, where h, b and c are arbitrary functions, refers to the transformed outcome functions of the problem (which result after a change of integration variable in (4.17)). We conclude that if one wants to allow both arbitrary utility functions and distribution functions, then the only outcome function yielding efficient risk-sharing is the constant outcome function!

The Need for Bounded Sharing Rules

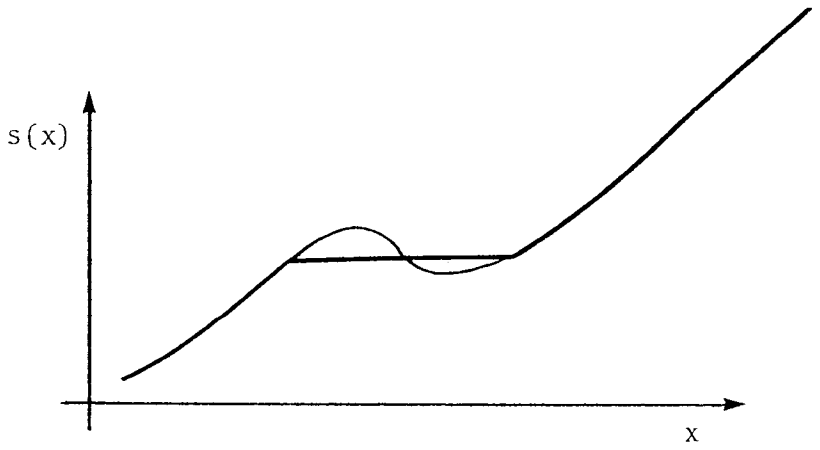
From (4.30) we can see why existence of a solution is a serious issue. Take for instance the Normal density function with mean = a and variance = 1:

$$f(x,a) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2}} .$$

Then $f_a/f = (x - a)$. This means that if $\mu \neq 0$ the RHS in (4.30) becomes negative for some x-values, whereas the LHS is always positive. Hence, $\mu = 0$ is the only possibility. But this cannot be true either, since then $s^*(x)$ provides optimal risk-sharing, which can easily be shown to imply improper action incentives when $V'(a) > 0$. The solution to the paradox, pointed out by Mirrlees [1974], is that there exists no optimal sharing rule in the class of unbounded functions. In fact, the first-best solution can in this case be approximated arbitrarily closely as we will see in Section 4.3.5. This motivated our restriction to bounded sharing rules, and the rigorous study of existence as well as of a characterization of the optimal sharing rule.

Characterizations in S_1 and S_2

(4.30) was derived for bounded measurable sharing rules. Let us briefly examine what happens if we optimize in S_1 or S_2 for which we have the existence results. A problem arises with the admissibility of the variation $s_h(x)$ in (4.26). There may be directions of increase of the Lagrangian, but such that all these directions take us outside the admissible class of functions (S_1 or S_2). Consequently, Theorem 4.7 will not hold in general. More complicated expressions than (4.30) could be derived. They would essentially say that either we have (4.30), or the class constraint is binding. For instance, an optimal non-decreasing function would follow the point-wise optimum along increasing parts and make jumps over decreasing parts (see picture below).



(for each μ there would generally exist a unique level at which the jump occurs.) From this it follows that if the point-wise optimum is nondecreasing, the characterization in corollary 4.8 is correct.[6] An important case for which this is true is recorded in the following:

Theorem 4.9: If s^* is optimal in the class of nondecreasing, bounded sharing rules S_1, and:

(i) $F_a(x,a^*) \leq 0$, for every x; $F_a(x,a^*) < 0$, for some x.

(ii) f_a/f is nondecreasing in x,

then s^* satisfies the optimality conditions in Corollary 4.8.

Proof: Let μ^* be a solution to (4.27) corresponding to the optimal solution (a^*,s^*). (4.27) can be written as:

$$E_a^P(a^*,s^*) + \mu^* \cdot E_{aa}^A(a^*,s^*) = 0, \tag{4.32}$$

since $E_a^A(a^*,s^*) = 0$. We also have:

$$E_{aa}^A(a^*,s^*) \leq 0. \tag{4.33}$$

We claim that $\mu^* > 0$. Suppose not, and study the two other possibilities: $\mu^* = 0$, $\mu^* < 0$.

Case I: $\mu^* = 0$. It follows that the point-wise optimum $\bar{s}(x)$ of the Hamiltonian is nondecreasing, since it provides optimal risk-spreading (see Wilson [1968]). Take

$$s_h(x) = h \cdot \bar{s}(x) + (1 - h) \cdot s^*(x), \quad \text{for all } x.$$

$s_h \varepsilon S_1$ for $h \geq 0$. Using the same argument as in Theorem 4.7, s_h provides a feasible direction of increase of $J(\cdot)$, unless $s^*(x) = \bar{s}(x)$. Since $s^*(x) = \bar{s}(x)$, the principal's share is strictly increasing (see Wilson [1968]), which in conjunction with assumption (i) implies $E_a^P(a^*,s^*) > 0$ (by first-order stochastic dominance). This contradicts $\mu^* = 0$ by (4.32) and (4.33).

Case II: $\mu^* < 0$. Let $\bar{r}(x) = x - \bar{s}(x)$ be the principal's share for the point-wise optimal sharing rule. $\mu^* \cdot \frac{f_a}{f}$ is nonincreasing by assumption (ii). On the other hand,

$$\frac{G'(\bar{r}(x))}{U'(x - \bar{r}(x))},$$

is increasing in x. Hence, for the characterization (4.30) to be valid, $\bar{r}(x)$ must be increasing. This implies that $x - s^*(x) = r^*(x) = \bar{r}(x)$ is also increasing, since $s^*(x)$ is either flat or follows $\bar{s}(x)$ (see picture above; to prove this rigorously use again a convex combination as a variation). We conclude, as in Case I, that $E_a^P(a^*,s^*) > 0$, contradicting $\mu^* < 0$ by (4.32) and (4.33).

We have shown that $\mu^* > 0$. From the fact that $G'(x - s(x))/U'(s(x))$ is decreasing in x, and assumption (ii), the characterization of Corollary 4.8 gives a nondecreasing point-wise optimum. Consequently, s^* must be point-wise optimal, which is the claim.

Q.E.D.

In the theorem above, assumption (i) is natural when action corresponds to some productive input like effort. From the next section on, we will work exclusively with this assumption. It is also true that f_a/f is increasing for many production functions when $\tilde{z}$ has some standard unimodal distribution (see transformations in Appendix 4A). For practical reasons one may furthermore want to restrict attention to nondecreasing sharing rules. Under such circumstances, Theorem 4.9 provides a characterization of an optimal sharing rule, which we also know exists by Theorem 4.5.

4.3.4 Properties of the Optimal Solution

The Second-Best Nature of the Solution

From now on we will assume:

B7 $U(w,a) = U(w) - V(a)$, $V'(a) > 0$, $V''(a) \leq 0$; $F_a(x,a) \leq 0$ for all x, and $F_a(x,a) < 0$ for some x, for all $a \in A$.

When B7 holds we will talk about $\underline{a}$ as effort. Let (a^*,s^*) denote an optimal solution, and μ^* the corresponding Lagrangian multiplier in (4.27). We have:

Lemma 4.11: $\mu^* \neq 0$; $E_a^P(a^*,s^*) \neq 0$.

Proof: $\mu^* \neq 0$ follows from the proof of Case I in Theorem 4.9, since we only used assumption B7 there. Suppose $E_a^P(a^*,s^*) = 0$. Then $\mu^* = 0$ is feasible from (4.27), contradicting the first part.

Q.E.D.

From the lemma follows:

Theorem 4.11: Assume B2-B7. Then there exists an action $a \varepsilon A$ and sharing rules s_1, $s_2 \varepsilon S_3$ such that both (a,s_1) and (a^*,s_2) are strictly Pareto superior to (a^*,s^*).

Proof: Define the feasible variation:

$$s_t(x) = s^*(x) + t, \text{ when } x \varepsilon X = \{x | s^*(x) < d\},$$

$$= s^*(x) \quad , \text{ when } x \varepsilon X^c.$$

Let $e^i(a,t) = E^i(a,s_t(x))$, which is a mapping from $\mathbb{R}^2$ to $\mathbb{R}$, for $i = A, P$. The set X must have positive measure, or else the agent would take an action on the boundary of A contradicting B7. This implies that the gradients of e^i are linearly independent at $a = a^*$, $t = 0$, since $E_a^P(a^*,s^*) \neq 0$ by the lemma, and $E_a^A(a^*,s^*) = 0$ by B7. Hence, there must exist a direction of strict increase for both e^A and e^P. $e^A(a^*,0) = 0$ implies that t must increase in this direction (the agent must receive money), so that $s_t(x) \varepsilon S_3$ for small changes. This proves the existence of (a,s_1).

Let $h(x) = f_a(x,a^*)/f(x,a^*)$. We claim $h(x) \neq 0$ on a set of

positive measure. Suppose not, i.e., $h(x) = k = \text{const. a.e.}$ Then we would have:

$$0 = \int f_a(x,a^*)dx = \int h(x)f(x,a^*)dx = k \cdot \int f(x,a^*) = k.$$

But $k = 0$ implies $f_a \equiv 0$ violating B7. We have $\mu^* \neq 0$ by the lemma, and so $h(x)$ nonconstant on a set of positive measure implies that s^* differs from the optimal risk-sharing rule $\bar{s}(x)$ on a set of positive measure (by (4.30)). Using Wilson's characterization of optimal risk-sharing (Wilson [1968]), s^* must then be inefficient. This establishes the existence of (a^*,s_2).

Q.E.D.

Theorem 4.11 embodies the second-best nature of (a^*,s^*). (a^*,s^*) is a solution which trades off risk-spreading advantages for proper effort incentives, without being optimal w.r.t. either one objective alone. As an immediate corollary we have the main theorem by Harris and Raviv [1976], extended to measurable sharing rules (as opposed to differentiable ones).

Corollary 4.12: Under assumptions B2-B7, there are returns to being able to observe and enforce the action.

Proof: When an action can be enforced, (a,s_1) of Theorem 4.11 can be attained which is Pareto superior to (a^*,s^*).

Q.E.D.

Remark: In Example 4.1 we saw that the first-best solution could be obtained, so Corollary 4.12 is generally false. The critical assumption, which does not hold in Example 4.1 is the existence of $f_a(x,a)$ for all x.

Undersupply of Effort

It seems natural to conjecture that $E_a^P(a^*,s^*) > 0$, when B7 holds. The reason is that the agent derives direct disutility from effort, whereas the principal does not. Hence, one would think that the principal would always prefer more effort than the agent provides when moral hazard is present. This is, of course, true if the principal's share is nondecreasing as we have argued earlier, but there is no guarantee that such will be the case at an optimum unless we make further assumptions. One sufficient condition is given by the following theorem:

Theorem 4.13: $\mu^* > 0$, or equivalently $E_a^P(a^*,s^*) > 0$, if $X_+ = \{x \in \mathbb{R} | f_a(x,a^*) \geq 0\} = [b,\infty)$, for some constant b.

Proof: The equivalence follows from (4.32)-(4.33) when we observe that $E_a^P(a^*,s^*) \neq 0$ by Lemma 4.11. By the same lemma we only have to show that $\mu^* < 0$ must be false.

Suppose $\mu^* < 0$. Let $r^*(x) = x - s^*(x)$ be the principal's share. We claim $\mu^* < 0$ implies $r^*(x_+) \geq r^*(x_-)$ for every pair $(x_+,x_-) \in X_+ \times X_-$, where $X_- = \{x | f_a(x,a^*) < 0\}$.

Pick $x_- \in X_-$ arbitrary. Two cases are possible in Corollary 4.8:

(i) $s^*(x_-) = d$. Since $s^*(x_+) \leq d$ and $x_+ > x_-$ we have $r^*(x_+) \geq r^*(x_-)$.

(ii) $s^*(x_-) < d$; $r^*(x_-) = x_- - c$. We have:

$$(4.34) \qquad 0 \leq G'(r^*(x_-)) - [\lambda + \mu^* \cdot \frac{f_a(x_-,a^*)}{f(x_-,a^*)}] \cdot U'(x_- - r^*(x_-))$$

$$< G'(r^*(x_-)) - [\lambda + \mu^* \cdot \frac{f_a(x_+,a^*)}{f_a(x_+,a^*)}] \cdot U'(x_+ - r^*(x_-)),$$

since $x_+ > x_-$, $f_a(x_-,a^*) < f_a(x_+,a^*)$, $U' > 0$, $U'' < 0$, and by assumption $\mu^* < 0$. We have earlier argued that the Hamiltonian H was concave or monotone, so (4.34) implies that r has to be increased for pointwise optimality. An increase is feasible w.r.t. to the constraint $s(x) \in [c,d]$, since $x_+ > x_-$. Hence, $r^*(x_+) > r^*(x_-)$. Cases (i) and (ii) establish the claim $r^*(x_+) \geq r^*(x_-)$ for every pair $(x_+,x_-) \in X_+ \times X_-$. It follows that $G(r^*(x_+)) \geq G(r^*(x_-))$. Thus,

$$(4.35) \qquad E_a^P(a^*,s^*) = \int_{X_+} G(r^*(x)) f_a(x,a^*)dx + \int_{X_-} G(r^*(x)) f_a(x,a^*)dx$$

$$\geq \int_{X_+} M \cdot f_a(x,a^*)dx + \int_{X_-} M \cdot f_a(x,a^*)dx$$

$$= M \cdot \int f_a(x,a^*)dx = 0.$$

Here M is such that $G(r^*(x_+)) \geq M \geq G(r^*(x_-))$ for every (x_+,x_-). Such an M exists by the previous argument. (4.35) together with Lemma

4.11 implies $E_a^P(a^*,s^*) > 0$, contradicting (4.32)-(4.33). Hence, $\mu^* \leq 0$ is false and the claim follows.

Q.E.D.

The following line of reasoning could provide a more general proof of $\mu^* > 0$. If $\mu^* < 0$, then $x \in X_+ \Rightarrow s^*(x) < \bar{s}(x)$ and $x \in X_- \Rightarrow s^*(x) > \bar{s}(x)$, where $\bar{s}(x)$ is the optimal risk-sharing rule. Now $B(a^*,s^*) < B(a^*,\bar{s})$. Let $\bar{a}$ be the agent's choice when $\bar{s}$ is employed. The agent prefers $(\bar{a},\bar{s})$ to $(a^*,\bar{s})$ by definition. If $\bar{a} \geq a^*$, so does the principal, since $x - \bar{s}(x)$ is nondecreasing and $F_a \leq 0$. Hence, $B(a^*,\bar{s}) \leq B(\bar{a},\bar{s})$ provided $\bar{a} \geq a^*$. This we have not been able to establish unless one makes assumptions similar to those in Theorem 4.13.

Properties of s^*

As we argued earlier it is true for many distributions that f_a/f is nondecreasing, and this, of course, is sufficient to ensure the assumption in Theorem 4.13. As a further characterization of $s^*(x)$ in case f_a/f is nondecreasing, we record the following corollary, which follows directly from Corollary 4.8 and Theorem 4.13.

Corollary 4.14: If f_a/f is nondecreasing in x, then $s^*(x)$ is a nondecreasing function of x.

As illustrations of these characterizations, let us look at some examples:

Example 4.2: Let

$$U(w,a) = U(w) - V(a) = 2\sqrt{w} - a^2$$

$$G(w) = w$$

$$f(x,a) = \frac{1}{a} \cdot e^{-\frac{k}{a}}, \quad x \geq 0.$$

The agent determines the mean of the distribution x by his choice of a. An example would be a repairman, whose effort determines the expected length of time the repaired machine will run before breaking down. The monetary outcome is proportional to the lifetime of the machine.

We notice that $f_a/f = \frac{1}{a^2}(x - a)$, which is increasing in x. We derive first s* without any constraints. Employing (4.30) we have:

$$s^*(x) = [\lambda + \mu \cdot \frac{(x - a)}{a^2}]^2. \tag{4.36}$$

The agent's first-order condition gives $\mu = a^3$.

Checking the second-order condition, shows that this μ-value corresponds to a maximum. Using (4.32) we get:

$$4a^3 + 2\lambda \cdot a = 1. \tag{4.37}$$

The first-best solution is easily seen to be:

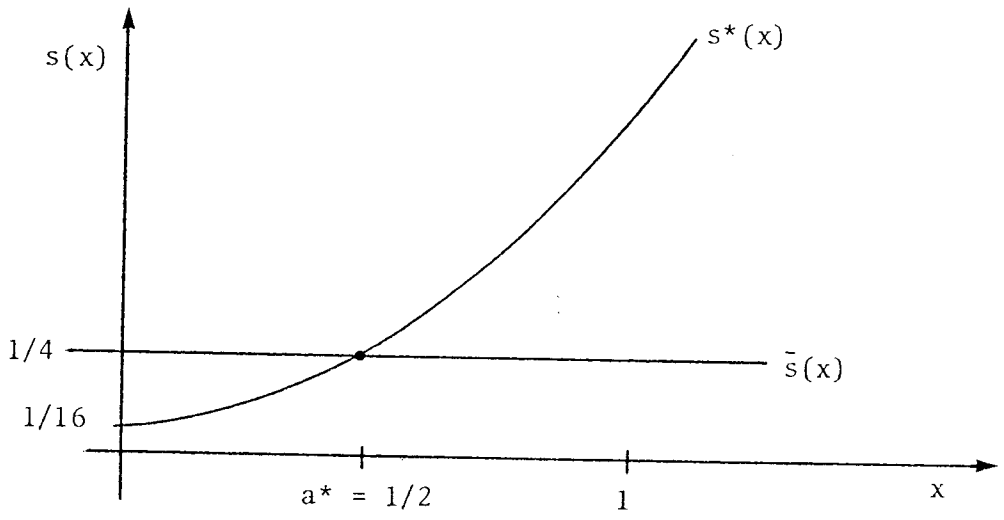
$$\bar{s}(x) = \lambda^2, \quad \bar{a} = \frac{1}{2\lambda}.$$

For a numerical solution, let $\lambda = \frac{1}{2}$. Substitution into (4.36) and (4.37) gives $a^* = \frac{1}{2}$, $\mu^* = \frac{1}{8}$, and

$$s^*(x) = \frac{1}{4}(x + \frac{1}{2})^2.$$

This gives $\frac{1}{2}$ as the welfare measure for the second-best solution compared to $\frac{3}{4}$ for the first-best solution.

The optimal functions are depicted below:



From the analysis we see that no lower bound needs to be imposed on s^*. However, an upper bound should have been used. But one can show that as the upper bound is raised, $\mu^* \to \frac{1}{8}$ and the solution converges to the one given above.

Reaffirming our earlier interpretation of (4.30), the agent is given incentives to increase his effort by penalties when $x \leq \frac{1}{2}$

and rewards when $x > \frac{1}{2}$. It is somewhat surprising to find the $s^*(x)$ is convex, considering risk-spreading, but this appears to be nothing uncommon when one looks at other examples. (Convex sharing rules are also familiar from Wilson's work on incentives in decentralized decision-making; Wilson [1969].) □

Example 4.3

$$U(w,a) = \log w - a^2$$

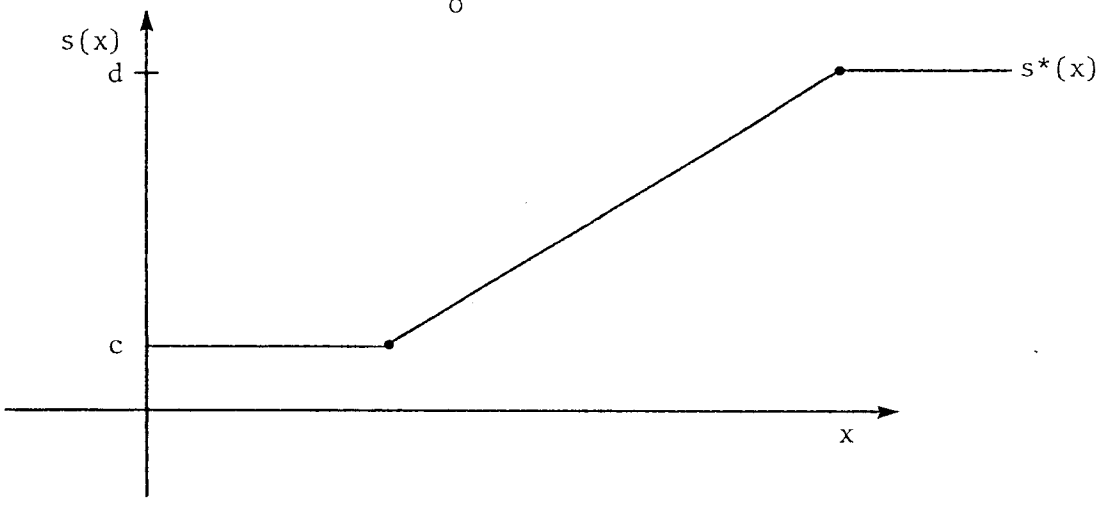
$$G(w) = w$$

$$f(x,a) = \frac{1}{\sqrt{2\bar{n}}\,\sigma} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} .$$

This distribution follows from the production function $x(a,z) = a + z$ with $\tilde{z} \sim w(0,\sigma^2)$.

In this example we have to impose upper and lower bounds explicitly. The solution is a piece-wise linear function:

$$s^*(x) = \lambda + \frac{\mu}{\sigma^2} \cdot (x - a^*), \quad c \leq s^*(x) \leq d.$$



It would be interesting to see how $s^*(x)$ changes with decreased variance. As the variance decreases we come closer to an ordinary contract where the agent is paid only if he supplies the agreed amount of effort (i.e., $s^*(x)$ is a step function). From the formula for s^* it looks as if a decrease in σ indeed makes the slope steeper, but we cannot be quite sure since μ also depends on σ. This dependence is hidden in the complicated joint solution of s^* and μ^* from (4.30) and (4.32), and is not easy to determine. □

By taking $U(w) = -e^{-w}$ in the previous example, we see that a concave sharing rule may be optimal, too. Sharing rules with decreasing parts can also be generated by using two-peaked distribution functions.

4.3.5 Approximations to the First-Best Solution

We know that (4.30) cannot hold for all x if $\mu \cdot f_a/f \to -\infty$ for some x-sequence. Mirrlees [1974] argued that in such cases the first-best solution can be approximated arbitrarily closely by a function, which uses severe penalties in very unlikely events. We will present Mirrlees argument with some modifications, and also discuss whether high bonuses will work as well for approximations.

Suppose $f_a/f \to -\infty$ as $x \to -\infty$ (this is the case in Example 4.3). Let $(\bar{a},\bar{s})$ be the first-best optimum. Approximate this solution by the following sharing rule:

$$(4.38)\qquad \begin{cases} s(x) = \bar{s}(x), & \text{if } x \geq g, \\ s(x) = \hat{s}(x;g), & \text{if } x < g. \end{cases}$$

Here $\hat{s}(x;g)$ is determined so that:

$$(4.39)\qquad \begin{cases} \text{(i)} & U(\hat{s}(x;g)) - U(\bar{s}(x)) = K(g), \quad \text{for all } x < g, \\ \text{(ii)} & \int_{-\infty}^{g} \{U(\hat{s}(x;g)) - U(\bar{s}(x))\} f_a(x,\bar{a})dx = K(g) \cdot F_a(g,\bar{a}) \\ & = V'(\bar{a}) - \int_{-\infty}^{\infty} U(\bar{s}(x)) f_a(x,\bar{a})dx = -E_a^A(\bar{a},\bar{s}) \equiv M. \end{cases}$$

(This is possible if U is not bounded from below.)[7] By this choice the agent's first-order condition for optimal effort is satisfied at $a = \bar{a}$. $E_a^P(\bar{a},\bar{s}) > 0$ by optimality of $\bar{s}$. Consequently, $E_a^A(\bar{a},\bar{s}) < 0$ and $M > 0$. $F_a < 0$ implies $K(g) < 0$ and hence $\hat{s}(s;g) < \bar{s}(x)$. From this it follows that the principal prefers $s(x)$ to $\bar{s}(x)$. Thus we only have to check that the agent's disutility from $s(x)$ compared to $\bar{s}(x)$ can be made negligible. This disutility is $K(g) \cdot F(g,\bar{a})$ for which we have:

$$(4.40)\qquad 0 \geq K(g) \cdot F(g,\bar{a}) = M \cdot \frac{F(g,\bar{a})}{F_a(g,\bar{a})} \qquad \text{(by (4.39))}.$$

But $f_a/f \to -\infty$ implies that for any $L < 0$, there is a $g(L)$ s.t. $f_a(g,x) \leq f(g,x) \cdot L$ when $g \leq g(L)$. Integration yields $F_a(g,x) \leq F(g,x) \cdot L$, implying $F/F_a \to 0$ as $g \to -\infty$. So we conclude that $K(g) \cdot F(g,\bar{a})$ can be made arbitrarily small.

We would still have to check the second-order condition:

$$E^A_{aa}(\bar{a},s) = K(g) \cdot F_{aa}(g,\bar{a}) + E^A_{aa}(\bar{a},\bar{s}) < 0?$$

But we do not really know the sign of $E^A_{aa}(\bar{a},\bar{s})$; only that $E^P_{aa}(\bar{a},\bar{s}) + \lambda \cdot E^A_{aa}(\bar{a},\bar{s}) \leq 0$. For most distributions $F_{aa}(g,\bar{a}) < 0$ for small g, and $F_{aa} \to 0$ as $g \to -\infty$ (see Appendix 4A), so we need $E^A_{aa}(\bar{a},\bar{s}) < 0$. This is true if the principal is risk-neutral, since then $\bar{s}$ is constant, but in general it depends on the form of the distribution of x. If the production function x(a,z) is concave in $\underline{a}$ and the sharing rule is concave (in addition to being increasing, which we know from the fact that it is efficient), then $E^A(a,\bar{s})$ is concave in $\underline{a}$, and the second-order condition is satisfied.

We can compare the approximation result to Example 4.1, where $\tilde{z}$ had compact support. In an imprecise sense $f_a/f = -\infty$ at the lower endpoint of the uniform distribution, in that example. This corresponds to our assumption that $f_a/f \to -\infty$, only that here $-\infty$ is "attained," which makes it possible to actually achieve the first-best solution. We see that even though the assumption of compact support may be artificial, the example carried substantial insight about the nature of the solution.

One may wonder if approximations to the first-best solution also can be achieved by high bonuses for exceptionally high outcomes, when $f_a/f \to +\infty$. The answer is no if the agent's utility function is bounded from above. This is easily seen since the first-order constraint

corresponding to (4.39) cannot be made to hold for arbitrarily high g-values because of the boundedness. A loss in incentives due to an increase in g cannot be compensated by an increase in the bonus, once g reaches a certain level.

On the other hand, if the agent's utility function has a linear asymptote and the principal is risk-neutral, then we can essentially reproduce the earlier argument to conclude that first-best approximations via high but unlikely bonuses is possible. We observe that the principal's disutility from the bonus can be approximated by a constant times the agent's utility when g is large, and this expression goes to zero when $g \to \infty$. Since the principal is risk-neutral, $E^A_{aa}(\bar{a},\bar{s}) < 0$, giving the correct second order constraint if $F_{aa} \to 0$.

What happens inbetween these two extreme cases (U bounded vs. U asymptotically linear) we do not know. If we look at Example 4.2 we find that the first-best solution cannot be attained by high bonuses (explaining our nonconcern for the upper bound), and it may generally be true that an asymptotically linear utility function by the agent is needed. If the principal is moreover risk-averse, then it is possible that even this will be insufficient. We notice that in Example 4.1 $f_a/f = +\infty$ (in an imprecise sense), but bonuses will, of course, not work. The reason is that here the second-order constraint will be violated.

Some qualitative conclusions emerge from the discussion above. We find that under certain circumstances the moral hazard problem can be essentially avoided using simple penalty schemes, provided there

are sufficient penalties available. Penalties seem to work more often than bonuses because of the concavity of the agent's utility function. Furthermore, moderate penalties suffice when the distribution is sufficiently sensitive to changes in effort; for instance, if the variance of the state of nature is small. From Corollary 4.12 we know that there are always returns to monitoring, but as we see, these returns may be negligible.

These conclusions seem to find some empirical support. We see quite often in practice simple dichotomous contracts of the form (4.38) (e.g., step functions). Maybe the threat of being fired can be considered one extreme example.

4.3.6 Generalized Distribution Function

In the previous derivations we have assumed all along that $F(x,a)$ has a density function for all $\underline{a}$. This assumption can be relaxed. We can allow a distribution which can be separated into two parts: one represented by a density function $f(x,a)$ as before, the other by a countable number of mass points $x_1, x_2, \ldots$, each with mass $f(x_i,a)$, $i = 1, 2, \ldots$. Notice that the positions of the mass points are assumed to be fixed, but the $f(x_i,a)$ may change with $\underline{a}$.

In many applications we find distributions with the above characteristics, particularly when they arise as compositions of several distributions. A good example is provided by any form of accident insurance. First, there is a probability of either having or not having an accident, and secondly, when an accident occurs there is a damage distribution.

Looking back at the arguments that were used for proving existence of an optimal s*, we find that little needs to be changed. The sequential compactness is immediate, since there are a countable number of mass points. The u.s.c. of J(s) does not change either, since we can still interchange the order of integration and limits using bounded convergence. In deriving a characterization of s*(x) we can also make the same arguments as before and arrive at the main formula (4.30), if we assume existence of $f_a(x_i,a)$ and $f_{aa}(x_i,a)$ for $i = 1, 2, \ldots$. Likewise, the conclusions about the second-best nature of an optimal solution are still valid.

It is interesting to note a special application of (4.30) to accident insurance. Suppose the agent's action only affects the probability of an accident, but not the extent of the damages. If the insurance company is risk-neutral, (4.30) tells us that the optimal insurance plan is a simple step function. In case of accident a deductible is paid, which is independent of the damage costs. To find the optimal deductible, we further need only know the mean of the damage distribution. This result may partly explain the frequent use of deductibles in health and accident insurance.

Notice that even though the agent is offered perfect insurance against the damage distribution, he is faced with risk resulting from the possibility of an accident. He has to pay a deductible, which serves as an incentive for action and hence the solution is inefficient by the standard of perfect information.

4.3.7 Additional Signals

In the preceding analysis x was the only observable. We will now look at extensions to situations where additional signals are observed. These signals could correspond to various kinds of monitoring of the agent, or to observations about the state of nature. Since we found that the solution to moral hazard was generally inefficient (compared to the standard of perfect information), one would expect that additional observations would be beneficial at least under certain circumstances.

Harris and Raviv [1976] study this issue, and in particular the question of when monitoring of the agent is valuable in the sense that including the signal outcome in the contract will make both parties strictly better off. Their results can be summarized in one sufficient condition. A signal is of value if:

(i) the agent can avoid any penalties with certainty by taking the agreed upon action,

(ii) the signal will detect any shirking with positive probability, and

(iii) there are sufficient penalties available.

Under these conditions the agent can be made to take the first-best action under first-best risk-sharing, as we saw in Example 4.1, and this is clearly sufficient for a Pareto improvement.

The conditions imposed on the signal by (i)-(iii) above are very strong, however, and one would expect that much less would suffice. But Harris and Raviv are quite skeptical about finding weaker conditions

which do not depend explicitly on the utility functions. In particular, they are concerned about relaxing (i), since then the risk-averse parties will face additional uncertainty introduced via the signal. They conjecture that such risk will occasionally outweigh the benefits from more information about the action.

We will show that such concerns are unwarranted. The main result is that a signal is valuable if and only if it provides information about the agent's action in addition to the information that one gets from observing the outcome x. This necessary and sufficient condition, which is only a property of the signal and its relation to the outcome x and action $\underline{a}$, completely solves the issue of when an additional information system has potential gains.

The result will follow from extending our characterization of efficient sharing rules to include signals. Let there be n signals observed besides x. Denote the signal vector by $y = (y_1, \ldots, y_n) \in \mathbb{R}^M$. Assume there is a joint distribution function $F(x,y;a)$, which has a density function $f(x,y;a)$ twice differentiable in $\underline{a}$. We are interested in efficient sharing rules $s(x,y): \mathbb{R}^{n+1} \to \mathbb{R}$. Such sharing rules can be generated by solving the program:

$$(4.41)\begin{cases} \max\limits_{a,s(x)} \int \{G(x - s(x,y)) + \lambda \cdot U(s(x,y),a)\} f(x,y;a)\,dxdy, \\ \text{s.t.} \quad \int \{U(s(x,y),a) f_a(x,y;a) + U_a(s(x,y),a) \cdot f(x,y;a)\}\,dxdy = 0. \end{cases}$$

To prove existence we have to restrict ourselves to bounded

sharing rules, which belong either to a family of equi-continuous functions or a class of coordinate-wise nondecreasing functions. We will not reproduce the proofs from Appendix 4B. Following exactly the steps for proving the main characterization in Corollary 4.8, we get the natural extension.

Corollary 4.15: Assume B2-B6 (with obvious changes in notation). Let (a^*,s^*) be an optimal solution satisfying B6 and let the agent's utility function be separable. Then $s^*(x,y)$ will maximize the Hamiltonian point-wise almost everywhere, that is for a.e. $(x,y)\in\mathbb{R}^{n+1}$:

$$(4.42) \qquad \frac{G'(x - s^*(x,y))}{U'(s^*(x,y))} = \lambda + \mu^* \cdot \frac{f_a(x,y;a^*)}{f(x,y;a^*)} ,$$

if the equation has a solution $c \leq s^*(x,y) \leq d$, or otherwise:

$$(i) \qquad s^*(x,y) = c, \quad \text{if} \quad G'(x - c) - (\lambda + \mu^* \cdot \frac{f_a}{f}) \cdot U'(c) > 0,$$

$$(ii) \qquad s^*(x,y) = d, \quad \text{if} \quad G'(x - d) - (\lambda + \mu^* \cdot \frac{f_a}{f}) \cdot U'(d) < 0.$$

Condition (4.42) is intuitively appealing. It shows that the agent's responsibility should be small when his action explains little of the variation in x conditional on y, and large in the opposite case. At an extreme, if $f_a(x,y;a^*) = 0$ for every outcome x, when a particular y obtains, the agent should be offered full risk sharing under this

contingency. This again perfectly matches the old accounting principle that managers should be held responsible only for what they can control (beyond what is optimal from a risk-sharing point of view, of course).

(4.42) has two important implications. First, it tells that provisions for unexpected occurrences (described by the signal y) should be included in the contract if these can be observed by both parties. Thus contracts can be expected to be elaborate and contain a variety of special arrangements under different contingencies. Certainly, there is substantial empirical support for this conclusion. Not only do contracts tend to be detailed, but in addition a host of provisions are left implicit. Natural disasters, strikes, accidents, are generally sufficient grounds for relieving an agent from full responsibility and allowing a change in the contract, even if such provisions are not explicitly written down. In these situations usually the legal system provides protection and fair resolution.

Secondly, (4.42) shows that there are gains to creating additional information systems; most naturally in employment relationships. Supervision and detailed performance measurement, particularly as developed in modern responsibility accounting (see Horngren [1972]), are examples of such information systems, and their role is explained by (4.42). Notice that these control systems are not only to the benefit of the employer, but also to the employee, who gets protected against events outside his control.

To illustrate how gains can be derived from additional monitoring, we can look at the following schematic example.

Example 4.4: The outcome x can be either good (=1) or bad (=0). This depends both on the action of the agent and the state of nature $\tilde{z}$. $z = (y_1, y_2)$, where y_1 and y_2 are independent. Each variable can take on values 0 or 1. The consequences and probabilities are as follows:

$$x = 1, \quad \text{if} \quad y_1 \cdot y_2 = 1,$$

$$x = 0, \quad \text{otherwise.}$$

$$P(y_1 = 1) = \frac{1}{2}, \quad P(y_2 = 1) = p(a), \quad 0 \leq p(a) \leq 1.$$

p(a) is an increasing function of the agent's action a. We have the following table of joint probabilities.

	$y_2 = 0$	$y_2 = 1$
$y_1 = 0$	$\frac{1}{2}$	0
$y_1 = 1$	$\frac{1 - p(a)}{2}$	$\frac{p(a)}{2}$

If $y_1 = 0$, the outcome will be bad regardless of the agent's action. Only in state $y_1 = 1$ can the agent influence the outcome. Let us first compare the cost (in terms of risk-sharing losses) of providing incentives for a fixed action when y_1 is observed and when it is not observed. A simple calculation shows that the following two schemes yield the same action:

I $\quad s(x) = w, \quad$ if $\quad x = 1,$

$\quad s(x) = v, \quad$ if $\quad x = 0.$

II $\quad s(x,y_1) = w, \quad$ if $\quad x = 1,\ y_1 = 1,$

$\quad = v, \quad$ if $\quad x = 0,\ y_1 = 1,$

$\quad = s(x,0), \quad$ if $\quad y_1 = 0;\ \ s(x,0)$ arbitrary.

The point is that $s(x,0)$ can be chosen based on mere risk spreading advantages. If the principal is risk-neutral it should be constant. The agent is freed from all responsibility when $y_1 = 0$, since nothing can be inferred about his action in this event. Since both schemes provide the same action incentives, but the second one provides superior risk-sharing, we see that observing y_1 is of value.

This value can also be seen from another viewpoint. Let $\underline{a}$ be the agent's action under scheme I. Marginally both the principal and the agent will be indifferent between a change in v such that:

$$\Delta v_0 \cdot \frac{1}{2} + \Delta v_1 \cdot \frac{1 - p(a)}{2} = 0,$$

where Δv_0 is the change in event $y_1 = 0$ and Δv_1 is the change when $y_1 = 1$. However, the incentives for action will change according to:

$$\Delta v_0 \cdot U'(v) \cdot 0 + \Delta v_1 \cdot U'(v) \cdot \frac{-p'(a)}{2} \gtrless 0. \tag{4.43}$$

The agent's utility will not change marginally from such a change in the action, since $\underline{a}$ was optimal, but the principal would like either an increase or decrease, unless we are at a first-best solution. By choosing the sign of Δv_1 correctly, the total result will be that risk-sharing returns stay intact but the action is improved when y_1 is introduced in the contract, and this leads again to an overall improvement. □

To summarize the results of the example: observing y_1 is valuable, since we can either get the same action with less costs in terms of lost risk-sharing opportunities, or get an improved action with original risk-sharing benefits. Normally, a mixture of both advantages will produce the new optimum (when y_1 is introduced) and this is exactly the content of condition (4.42). We notice that such advantages can be achieved for each outcome of y separately, and so integrating over y will result in a total gain. This explains why we need not be concerned about the noisiness of the signal $\tilde{y}$ as Harris and Raviv [1976] were.

(4.42) suggests that a contract $s(x)$ can be dominated by a contract $s(x,y)$ if and only if it is <u>false</u> that:

$$(4.44) \qquad \frac{f_a(x,y;a^*)}{f(x,y;a^*)} = \bar{h}(x;a^*), \quad \text{for} \quad \text{a.e. } (x,y)\in\mathbb{R}^{n+1} .$$

However, (4.42) does not quite prove this, since the program in (4.41) does not generally generate all efficient points, because the efficiency

frontier may not be concave (see footnote 3). In that sense (4.42) only implies that efficient points on the concave hull of the frontier are dependent of y when (4.44) is false.

There is also a second technical difficulty involved with the conjecture. (4.44) is a condition at a*. We can prove that if (4.44) is false at a*, then the efficient contract must depend on y. But the reverse cannot be proved, since from (4.44) we get no information about what happens outside a*. However, the cases for which (4.44) is true only at a* are not very interesting and certainly not of practical significance, and so we want to exclude them from our analysis.

This motivates the following slightly asymmetric

Definition: $\tilde{y}$ is said to be informative about $\underline{a}$ if for all $\underline{a}$ it is false that:

$$(4.45) \qquad \frac{f_a(x,y;a)}{f(x,y;a)} = \bar{h}(x;a), \quad \text{for} \quad \text{a.e. } (x,y)\in\mathbb{R}^{n+1} .$$

It is said to be noninformative about $\underline{a}$ if (4.45) is true for all $\underline{a}$.

As we said, these two cases are not perfect complements, but the exceptional cases are uninteresting and rare.

The content and meaning of (4.45) is not immediate. For a proper interpretation we can derive an equivalent condition which is more easily understood. Suppose $\tilde{y}$ is noninformative. Then we can solve (4.45) as a partial differential equation. The unique solution is:

$$f(x,y;a) = h(x;a)g(x,y), \tag{4.46}$$

where we can take $h,g \geq 0$. Conversely, (4.46) implies (4.45), since

$$f_a(x;a) = \int f_a(x,y;a)dy = h_a(x;a) \int g(x,y)dy,$$

$$f(x;a) = \int f(x,y;a)dy = h(x;a) \int g(x,y)dy,$$

where $f(x;a)$ is the marginal distribution of x.

If (4.46) is true for all $\underline{a}$ (for almost every $(x,y)\in\mathbb{R}^{n+1}$), then $\tilde{y}$ is noninformative. This has an intuitive explanation. (4.46) is precisely the condition for a sufficient statistic, if we interpret $\underline{a}$ as a random variable. It says that x is a sufficient statistic for the pair (x,y) w.r.t. a. In other words, x carries all the relevant information about $\underline{a}$, and y adds nothing to the power of inference. y could only be of value for risk sharing, but we know that optimal risk sharing is independent of the distribution of the random variable, when agents have homogeneous beliefs. Consequently, y should be valueless, which is what (4.42) indicates.

With these preliminaries we can prove the main theorem:

Theorem 4.16: If y is noninformative about $\underline{a}$, then an optimal sharing rule need not depend on y. If y is informative about $\underline{a}$, then an optimal sharing rule has to depend on y.

Proof: Suppose y is noninformative. Let $s(x,y)$ be an arbitrary sharing rule. We will show that there is a sharing rule $s(x)$ which is at least as good as $s(x,y)$, establishing the assertion that optimal sharing rules need only depend on x when y is noninformative.

For every x, define $s(x)$ so that

$$(4.47) \qquad \int U(s(x,y))g(x,y)dy = \int U(s(x))g(x,y)dy$$

$$= U(s(x)) \cdot \int g(x,y)dy = U(s(x)) \cdot k(x),$$

where $k(x) = \int g(x,y)dy$. Then using (4.46):

$$\int U(s(x,y))f_a(x,y;a)dxdy$$

$$= \int U(s(x,y))h_a(x;a)g(x,y)dxdy$$

$$= \int U(s(x))h_a(x;a)g(x,y)dxdy.$$

Similarly,

$$\int U(s(x,y))f(x,y;a)dxdy = \int U(s(x))f(x,y;a)dxdy.$$

Consequently, $s(x)$ will result in the same action and welfare for the agent.

(4.47) implies, by Jensen's inequality,

$$\int s(x,y)g(x,y)dy \geq \int s(x)g(x,y)dy,$$

or

$$\int (x - s(x,y))g(x,y)dy \leq \int (x - s(x))g(x,y)dy.$$

This implies, using Jensen's inequality a second time, that:

$$\int G(x - s(x,y))g(x,y)dy \leq \int G(x - s(x))g(x,y)dy.$$

Since this is true for every x, and $h(x;a) \geq 0$, we have by integrating,

$$\int G(x - s(x,y))f(x,y;a)dxdy \leq \int G(x - s(x))f(x,y;a)dxdy.$$

Since the agent takes the same act with s(x) as with s(x,y) by construction, this shows that the principal is at least as well off with s(x) as with s(x,y). The agent's utility is the same for both s(x) and s(x,y), and thus s(x) is weakly Pareto superior to s(x,y), which proves the first part of the theorem.

Let s(x) be arbitrary and fix x. The principal's and the agent's returns, conditional on x, from a variation $\delta s(x,y)$ in the sharing rule s(x) are:

(4.48)
$$\delta E^P = -G'(x - s(x)) \int \delta s(x,y)f(x,y;a)dy$$
$$+ \mu \cdot U'(s(x)) \int \delta s(x,y)f_a(x,y;a)dy,$$
$$\delta E^A = U'(s(x)) \int \delta s(x,y)f(x,y;a)dy.$$

Here μ is the Lagrangian multiplier in (4.27) corresponding to $s(x)$.

Suppose y is informative. Then there exist two sets Y and Y^c in the range of y, s.t.

$$(4.49) \qquad \frac{f_a(x,Y;a)}{f(x,Y;a)} \neq \frac{f_a(x,Y^c;a)}{f(x,Y^c;a)} ,$$

where $\int_Y f(x,y;a)dy = f(x,Y;a)$ and correspondingly for the other.

Choose a variation $\delta s(x,y)$ such that $\delta s(x,y) = 0$ for $y \notin Y \cup Y^c$, and

$$(4.50) \qquad \delta s(x,Y) \cdot f(x,Y;a) + \delta s(x,Y^c) f(x,Y^c;a) = 0$$

($\delta s(x,Y)$ is constant for all $y \in Y$ and correspondingly for $\delta s(x,Y^c)$).

From (4.48) and (4.50) it follows that:

$$\delta E^P = \mu \cdot U'(s(x)) [\delta s(x,Y) \cdot f_a(x,Y;a) + \delta s(x,Y^c) \cdot f_a(x,Y^c;a)],$$

and

$$\delta E^A = 0.$$

Substituting from (4.50):

$$\delta E^P = \mu \cdot U'(s(x)) \cdot \delta s(x,Y) \cdot f(x,Y;a) [\frac{f_a(x,Y;a)}{f(x,Y;a)} - \frac{f_a(x,Y^c;a)}{f(x,Y^c;a)}].$$

$\mu \neq 0$, $f \neq 0$, $U' > 0$ and the parenthesis is $\neq 0$. Hence $\delta s(x,Y)$ can be chosen so that $\delta E^A = 0$ and $\delta E^P > 0$. This is true for a fixed x, which has zero mass. But since y is informative, the same can be done for

a set of x-values with positive mass, implying that strict Pareto improvements can be made.

Q.E.D.

Remark: It is clear that Theorem 4.16 is also valid when F is a generalized distribution function as discussed in the previous section.

Theorem 4.16 has the following obvious corollary.

Corollary 4.17: If y is noninformative about a, then there are no returns to observing y. If y is informative about a, then there are positive returns to observing y.

Using the verbal interpretation provided by (4.46), we see that monitoring is of value if and only if y provides any information about a in addition to what one can infer from x. The fact that any extra information is valuable, regardless of how noisy it is, is the strong part of the theorem, and maybe somewhat surprising (compare with the earlier mentioned reasoning by Harris and Raviv). But it is easily understood in view of (4.42). The intuition of (4.42) is that whenever $\tilde{y}$ is informative, both parties can be made better off in each state of y, by including y in the contract. This results, of course, in overall improvement.

In our formulation we have made no distinction between the signal being an observation of the action or of the state of nature. Indeed, the analysis shows that such a dichotomy is not essential.

A duality exists such that it is almost a matter of semantics (from the theoretical point of view) which interpretation the signal is given. From a practical standpoint, there is, of course, a difference, and we will now look at the case where the signal is an independent observation about the action. In that case:

$$f(x,y;a) = h(x;a)g(y;a). \tag{4.51}$$

From this follows:

$$\frac{f_a(x,y;a)}{f(x,y;a)} = \frac{h_a(x;a)}{h(x;a)} + \frac{g_a(y;a)}{g(y;a)} .$$

Hence, y is noninformative if and only if:

$$\frac{g_a(y;a)}{g(y;a)} = k = \text{const.}, \qquad \forall y.$$

This implies,

$$0 = \int g_a(y;a)dy = k \cdot \int g(y;a)dy = k.$$

Consequently, $g_a(y;a) = 0$ for every y, which means that the distribution of y is independent of a.

Whenever y is such that the distribution depends on a, it is informative. A simple example is provided by the following monitoring technology. The principal makes a random check that the agent is

working when he is supposed to. This information in itself may not tell that the agent is doing something wrong -- he may just take a well-deserved break -- but it has the property that it is more likely that the agent will be on a break if he is not providing the proper effort. In other words, if we write $y = 0$ for the observation that the agent is on a break and $y = 1$ otherwise, we have that $P(y = 1|a)$ is an increasing function of $\underline{a}$. Thus y is informative about $\underline{a}$ and from Theorem 4.16 it follows that however weak the signal is, it has potential gains.

The positive value of this particular information system was proved in Gjesdal [1976]. It seems to correspond well to certain behavioral controls used in practice; e.g., labor supervision.

Theorem 4.16 tells nothing about how valuable the additional information will be. This would be important, since information is always costly to obtain. In our discussion of approximation results, we found that under certain circumstances the first-best solution can be approximated arbitrarily closely when sufficient penalties are available, and thus returns from additional monitoring become negligible. However, in practice penalties are restricted and if the variance of $\tilde{z}$ is relatively high, this restriction may substantially reduce efficiency when contracts only depend on the outcome x. Though we did not explicitly include bounds on the sharing rule in analyzing returns from monitoring, it is clear that an information system can play a crucial role in utilizing the available penalties more efficiently and improve on the solution.

Some indications of the value of additional signals can be found by studying (4.42). We see that the more variation in f_a/f one can achieve with a signal y, the more valuable it will be. Put in another way: the more variation the conditional likelihood function from y exhibits, the higher returns we can expect from it.

4.4 Concluding Remarks

We have discussed two aspects of moral hazard in this chapter: team production, and principal-agent contracts under uncertainty.

We found that team production resulted in inefficient supply of inputs when agents shared the total outcome and no additional information systems were available. The central problem was that the budget had to balance in cooperative institutions. We concluded that a natural development of the organization would be to separate ownership from input supply, since then the problems with team production could be avoided. This provides in our view one explanation for the emergence of corporations. It differs somewhat from Alchian and Demsetz's theory of the firm, since monitoring has a central position in their arguments.

Though monitoring can be seen to play some role in an environment of certainty, it gains additional significance when uncertainty is present. We studied uncertainty in the context of a principal-agent relationship. A characterization of efficient contractual agreements were obtained, both based on the outcome alone and when additional signals were available. In the former case, we emphasized

rigor by proving the existence of an optimal solution in a restricted class of sharing rules, and carefully deriving optimality conditions. This was motivated by the fact that important examples of nonexistence were known.

We also studied the approximation results of Mirrlees [1974] with the qualitative conclusion that in many instances moral hazard does not pose as severe problems as one may expect. Particularly, if the agent can be evaluated over a longer time span, the variance in judgment becomes small, and simple penalty or bonus schemes will become effective. In situations where this is not the case, additional information systems become essential. We gave a simple necessary and sufficient condition for such monitoring to be of value. The analysis provided also an explanation for the complex contractual agreements that are normally observed in practice.

So far the two kinds of moral hazard have been treated separately, but from the analysis it is possible to draw some conclusions about what happens in a mixed situation with n agents and state uncertainty. Certainly, budget-balancing is again an essential issue. If it is imposed, so that the team members who provide productive inputs also share the outcome, one can expect this arrangement to be less efficient than if separate ownership is established. Monitoring will in that case play a dual role (and accordingly be more important), since discerning individual actions will both make decoupling possible and allow for improved risk sharing. However, notice that this is not to say that each agent should only be responsible for what he can

control, since in that case one would generally forego opportunities for risk sharing among agents. For instance, in a partnership everybody may gain from sharing the outcomes of each others products (just as in the two-person principal-agent case), but ideally monitoring should prevent such an arrangement to lead to distortions in the supply of effort.

When budget-balancing can be relaxed, say via a separate ownership, the analysis becomes essentially similar to the two-person case, though there may again exist potential gains from risk sharing between agents (unless the principal is risk neutral). Because of the presence of uncertainty, there will not exist simple schemes which will guarantee efficiency, and monitoring will generally play an important role in insuring proper actions and improved risk sharing.

An extension to our analysis of moral hazard would take into account the possibility of differential information about state uncertainty between the principal and the agents. Quite often, the agent who supplies the productive input is also better aware about the difficulty of his task. In that case a second-best solution would generally require that the agent is given some freedom in determining the sharing rule. We have seen an example of this in the use of a goal-based incentive scheme (Example 2.4) and our analysis of management by participation (Example 2.7) gave some indications of how such schemes may work.

APPENDIX 4.A

SOME PARTICULAR OUTCOME FUNCTIONS

The following three outcome functions illustrate the relationship between the two alternative problem formulations.

I. $x(a,z) = z \cdot h(a); \quad z \geq 0, \quad z \sim \hat{F}; \quad h(a) > 0, \; h'(a) > 0, \; h''(a) < 0.$

$$F(x,a) = \hat{F}\left(\frac{x}{h(a)}\right)$$

$$F_a(x,a) = \hat{f}\left(\frac{x}{h(a)}\right) \cdot -\frac{h'(a)}{h^2(a)} < 0$$

$$\frac{F(x,a)}{F_a(x,a)} \sim -\frac{\hat{F}\left(\frac{x}{h(a)}\right)}{\hat{f}\left(\frac{x}{h(a)}\right)}$$ ($\sim$ means proportional to, when a const.)

$$f(x,a) = \hat{f}\left(\frac{x}{h(a)}\right) \cdot \frac{1}{h(a)}$$

$$f_a(x,a) = \hat{f}'\left(\frac{x}{h(a)}\right) \cdot \frac{h'(a)}{h(a)} + \hat{f}\left(\frac{x}{h(a)}\right) \cdot \frac{h'(a)}{h^2(a)}$$

II. $x(a,z) = z + h(a)$

$$F(x,a) = \hat{F}(x - h(a))$$

$$F_a(x,a) = \hat{f}(x - h(a))(-h'(a)) < 0$$

$$\frac{F(x,a)}{F_a(x,a)} \sim - \frac{\hat{F}(x - h(a))}{\hat{f}(x - h(a))}$$

$$f(x,a) = \hat{f}(x - h(a)) \qquad f_a(x,a) = \hat{f}'(x - h(a)) \cdot (-h'(a)).$$

III. $x(a,z) = h(a + z), \quad z \geq 0.$

$$F(x,a) = \hat{F}(h^{-1}(x) - a)$$

$$F_a(x,a) = -\hat{f}(h^{-1}(x) - a) < 0$$

$$\frac{F(x,a)}{F_a(x,a)} \sim - \frac{\hat{F}(h^{-1}(x) - a)}{\hat{f}(h^{-1}(x) - a)}$$

$$f(x,a) = \hat{f}(h^{-1}(x) - a) \cdot \frac{1}{h'(h^{-1}(x))} .$$

$$f_a(x,a) = \hat{f}'(h^{-1}(x) - a) \frac{-1}{h'(h^{-1}(x))}.$$

APPENDIX 4.B

PROOF OF THEOREM 4.5

Theorem 4.5 will be proved in a sequence of lemmata.

Lemma 4B.1: Let $\{s_n\}$ be a sequence of functions in S_1. Then there exists a subsequence $\{s_{n'}\}$ and a function $s \in S_1$ such that

$$\lim_{n' \to \infty} s_{n'}(x) = s(x), \qquad \text{for almost every } x \in \mathbb{R}.$$

Proof: The rational numbers can be ordered since they are countable. Let $\{x_m\}$ be a sequence containing all rational numbers. $s_n(x_1) \in [c,d]$ for every n, and consequently there is a convergent subsequence $\{s_{n^1}(x_1)\}$ with limit $s(x_1) \in [c,d]$. Likewise, $\{s_{n^1}(x_2)\}$ has a convergent subsequence $\{s_{n^2}(x_2)\}$, where n^2 is a refinement of n^1. Its limit is denoted $s(x_2) \in [c,d]$. Continue in this manner to define a function $s(x)$ on all rationals.

The construction yields a sequence of refinements $n^1 \geq n^2 \geq \ldots$ associated with the sequence $\{x_m\}$. Define a further refinement n' of n as follows: take the first element in n^1, the second in n^2 and so on. For this refinement it holds that, from its m^{th} member on, it is a refinement of n^m. Hence, $s_{n'}(x) \to s(x)$ for all rational x.

Extend $s(x)$ to all of $\mathbb{R}$ by defining $s(x) = \limsup s_{n'}(x)$,

for every $x \in \mathbb{R}$. $s(x)$ is easily seen to be nondecreasing and have range $[c,d]$, since this is true for each $s_{n'}$. Hence, $s \in S_2$. It also follows that the set X of points of discontinuity of s has measure zero. (A bounded nondecreasing function can have at most a countable number of discontinuities.)

Take any point $x \in X^c$. For any rational number $r < x$ we have,

$$s(r) = \liminf s_{n'}(r) \leq \liminf s_{n'}(x) \leq s(x).$$

The equality follows by construction of the sequence n' and the definition of s; the first inequality holds true since each $s_{n'}$ is nondecreasing, and the second by definition of s. Letting $r \to x$ we get, since s is continuous at x,

$$\liminf s_{n'}(x) = s(x), \quad \text{for every } x \in X^c,$$

which gives our claim by definition of s and the fact that X has Lebesque-measure zero. Q.E.D.

Lemma 4B.2: Let $\{s_n\}$ be a sequence of functions in S_2. Then there exists a subsequence $\{s_{n'}\}$ and a function $s \in S_2$ such that

$$\lim_{n' \to \infty} s_{n'}(x) = s(x) \qquad \text{for every } x \in \mathbb{R}.$$

Proof: Repeat the construction in the proof of the previous lemma to get a subsequence $\{s_{n'}\}$, which converges pointwise on all rationals. Define as before $s(x) = \limsup s_{n'}(x)$. We show first that $\lim_{n' \to \infty} s_{n'}(x) = s(x)$.

Let $x \in \mathbb{R}$ be arbitrary, and take y rational and such that $|x - y| < \delta(\varepsilon)$. Then,

$$|s_{n'}(x) - s_{n'}(y)| \leq \varepsilon \qquad \text{for every } n'.$$

Writing out this as a double inequality and taking lim sup's and lim inf's gives us:

$$\limsup s_{n'}(x) \leq s(y) + \varepsilon$$
$$\liminf s_{n'}(x) \geq s(y) - \varepsilon,$$

using the fact that $s_{n'}(y)$ converges, since y is rational. Since $\varepsilon > 0$ can be taken arbitrarily small we get that $s_{n'}(x)$ converges pointwise everywhere.

To show that $s \in S_2$ note first that it has range $[c,d]$. Let x,y be such that $|x - y| < \delta(\varepsilon)$, but otherwise arbitrary. Then,

$$|s_{n'}(x) - s_{n'}(y)| \leq \varepsilon \qquad \text{for every } n'.$$

Take the limit as $n' \to \infty$ to get

$$|s(x) - s(y)| \leq \varepsilon.$$

Hence, s belongs to the family of equi-continuous functions with modulus $\delta(\varepsilon)$. Q.E.D.

We have shown that for both classes S_i, $i = 1, 2$, we can find a.e. convergent subsequences of a sequence. Next we turn to the upper semi-continuity of the objective functional under the notion of a.e. convergence. (Note that S_1 and S_2 can be topologized so that a.e. convergence in S_1 and pointwise convergence in S_2 are the induced modes of convergence.) We begin by

Lemma 4B.3: The agent's solution correspondence a(s) is u.s.c. w.r.t convergence a.e.

Proof: We have to show that if

(i) $s_n \to s$ (a.e.),

(ii) $a_n \in a(s_n)$ for every n,

(iii) $a_n \to a$,

then $a \in a(s)$.

Write $M(a,s) = \int U(s(x),a)f(x,a)dx$. By B2 and B3,

$$U(s_n(x),a_n) \cdot f(x,a_n) \to U(s(x),a) \cdot f(x,a) \text{ for every } x.$$

By the fact that $s_n(x) \in [c,d]$ for every x, we can use the bounded convergence theorem to conclude that $M(a_n,s_n) \to M(a,s)$.

Let $a^* \in a(s)$. We need to show that $M(a,s) = M(a^*,s)$. We have $M(a,s) \leq M(a^*,s)$ by definition. For the other inequality we have:

$$M(a^*,s) \leq M(a^*,s_n) + \varepsilon \leq M(a_n,s_n) + \varepsilon \leq M(a,s) + 2\varepsilon,$$

when $n \geq n_0(\varepsilon)$, where $n_0(\varepsilon)$ is determined so that both the first and the last inequality holds based on continuity. The middle inequality follows by definition of a_n. Since $\varepsilon > 0$ can be taken arbitrarily small we conclude $M(a^*,s) \leq M(a,s)$, completing the proof. Q.E.D.

Recall the definitions:

$$B(a,s) = \int [G(x - s(x)) + \lambda \cdot U(s(x),a)] f(x,a)dx,$$

and

$$J(s) = B(a_{max}(s),s).$$

Lemma 4B.4: $J(s)$ is u.s.c. with respect to convergence almost everywhere.

Proof: We need to show that $s_n \to s^*$ a.e. implies $J(s_n) \to J \leq J(s^*)$. By compactness of A, $a_n \equiv a_{max}(s_n)$ has a convergent subsequence $a_{n'} \to a^* \in A$. $B(a_n,s_n) \to B(a^*,s^*)$ by B2, B3, B4 and by using

the bounded convergence theorem. By the previous lemma $a^* \varepsilon a(s^*)$ and so by definition of a_{max}, $J = B(a^*,s^*) \leq B(a_{max}(s^*),s^*) = J(s^*)$.

Q.E.D.

This lemma reveals the reason why we assumed B5.

We can now prove our main theorem:

Theorem 4.5: Let $S = S_1$ or S_2. Assume B1-B5. Then there exists an optimal solution to problem (4.25), (a^*,s^*), with $a^* = a_{max}(s^*)$.

Proof: Let $\bar{J} = \sup_{s \varepsilon S_i} J(s) < \infty$, i = 1 or 2. By definition there exists a sequence $\{s_n\}$ such that $\lim_{n \to \infty} J(s_n) \to \bar{J}$. By Lemma 4B.1 (or 4B.2 if i = 2) there exists a subsequence $\{s_{n'}\}$ which converges a.e. to an $s^* \varepsilon S_i$. Of course, $\lim_{n' \to \infty} J(s_{n'}) \to \bar{J}$. By Lemma 4B.4 $\lim_{n' \to \infty} J(s_{n'}) \leq J(s^*)$, and so $J(s^*) = \bar{J}$, implying the supremum is attained by $(a_{max}(s^*),s^*)$.

Q.E.D.

Footnotes to Chapter IV

[1]This assumption simplifies the analysis but is not essential for the results.

[2]In Kleindorfer and Sertel [1976a], it is shown that there exists a unique Nash equilibrium if f_i and x are strictly concave.

[3]Because of possible nonconvexities induced by the action $\underline{a}$ and constraint (4.18), it may not be possible to get all efficient pairs as solutions to the program.

[4]By a first-best solution we mean one which would be optimal under conditions of complete observability.

[5]For other purposes than ours, formulation (4.12)-(4.19) may be more appropriate. For instance when Stiglitz [1974, 1975] studies linear sharing rules it is easier to work with constraint (4.19) than with (4.21).

[6]But notice that this cannot be checked by solving μ and $s(x)$ from the characterization in Corollary 4.8 and conclude that $s(x)$ is optimal in S_1 if it is nondecreasing.

[7]Mirrlees uses another approximation rule, namely

$$s(x) = \bar{s}(x), \quad \text{if} \quad x \geq g,$$

$$s(x) = \varepsilon \quad , \quad \text{if} \quad x < g.$$

Our derivation is simpler because of our choice of approximation rule, but Mirrlees' scheme is easier to implement.

CHAPTER V

EPILOGUE

In this dissertation we have addressed problems of incentives in organizations, which arise due to asymmetric information among members of the organization. In the framework of a general game-theoretic formulation of the incentive problem, we have studied three specific topics: delegation, coordination of information, and the supply of productive inputs. The emphasis of the analysis has been theoretical, aimed at a better understanding of how the organization can develop various incentive or control schemes to overcome problems related to asymmetric information. Though we have not been looking for results of direct practical applicability, some of them are potentially useful. This is particularly true for the chapter on moral hazard. The model of team production is suggestive of organizational design and the structure of internal accounting. A more detailed analysis of the optimal sharing rule in the principal-agent model (or of some simpler contract form) should also prove useful for practice.

The scope of the dissertation is wide and many potentially fruitful directions for further research suggest themselves. A main issue of interest is the study of observability. We have seen that what is observable to the principal and the agents is the key factor

that determines what is achievable via cooperation. Being somewhat speculative, one might hope that a more basic theory of incentives could be developed by studying how the noncooperative game between members of the organization can be transformed by altering the condition of observability in order to yield more efficient solutions. An indication of the direction such an analysis could take, is provided in Moulin [1977], where "self-punishment" in a two-person game is studied.

Some more specific research topics include:

- A study of goal-based incentive schemes for control in a centrally planned economy or a firm (see Example 2.4), as well as other outcome-based incentive schemes.

- A characterization of environments for which strongly incentive compatible mechanisms exist.

- An analysis of the use of simple bonus or penalty schemes (i.e., step functions) in moral hazard problems.

- An analysis of the effects of asymmetric information in conjunction with moral hazard problems, and the use of goal-based incentive schemes as control mechanisms.

REFERENCES

Alchian, A., and H. Demsetz [1972], "Production, Information Costs, and Economic Organizations," American Economic Review, Vol. 62.

Arrow, K. [1965], "Aspects of a Theory of Risk-Bearing," Yrjö Jahnsson Lectures, Helsinki.

Barlow and Proschan [1972], Statistical Theory of Reliability and Life Testing, Holt, Rinehart and Winston, Inc.

Benassy, J. P. [1975], "Neo-Keynesian Disequilibrium: Theory in a Monetary Economy," Review of Economic Studies.

d'Aspremont, C., and L. A. Gerard-Varet [1975], "Individual Incentives and Collective Efficiency for an Externality Game with Incomplete Information." Discussion paper, CORE, July.

Gjesdal, F. [1976], "Accounting in Agencies," mimeo, Graduate School of Business, Stanford University, September.

Green, J., and J. J. Laffont [1978], Incentives in Public Decision Making, North-Holland, Amsterdam (forthcoming).

Green, J., and J. J. Laffont [1977], "Characterization of Strongly Individually Incentive Compatible Mechanisms for the Revelation of Preferences for Public Goods," Econometrica.

Grossman, S. [1976], "On the Efficiency of Competitive Stock Markets Where Traders Have Diverse Information," Journal of Finance, May.

Grossman, S., R. Kihlstrom, and L. Mirman [1977], "A Bayesian Approach to the Production of Information and Learning by Doing," Review of Economic Studies, October.

Groves, T. [1973], "Incentives in Teams," Econometrica, July.

Groves, T. [1974], "Information, Incentives and the Internalization of Production Externalities," The Center for Mathematical Studies in Economics and Management Science, Discussion Paper No. 87, Northwestern University.

Groves, T. [1975], "Incentive Compatible Control of Decentralized Organizations," Discussion Paper No. 166, Graduate School of Management, Northwestern University, August.

Groves, T., and M. Loeb [1975], "Incentives and Public Inputs," Journal of Public Economics, 4.

Groves, T., and J. Ledyard [1977], "Optimal Allocation of Public Goods: A Solution to the Free Rider Problem," Econometrica.

Harris, M., and A. Raviv [1976], "Optimal Incentive Contracts with Imperfect Information," Working Paper, Graduate School of Industrial Administration, Carnegie-Mellon University.

Harsanyi, J. [1967-1968], "Games of Incomplete Information Played by Bayesian Players," Parts I-III, Management Science, 14.

Heal, G. [1973], The Theory of Economic Planning, North Holland, Amsterdam.

Horngren, C. [1972], Cost Accounting: A Managerial Emphasis (3rd ed.), Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Hurwicz [1972], "On Informationally Decentralized Systems," in McGuire and R. Radner (eds.), Decision and Organization, North Holland, Amsterdam-London.

Hurwicz [1976], "Outcome Functions Yielding Walrasian and Lindahl Allocations at Nash Equilibrium Points," mimeo, University of Minnesota, November.

Ireland, N. [1977], "Ideal Prices vs. Prices vs. Quantities," Review of Economic Studies, February.

Jennergren, L. [1971], "Studies in the Mathematical Theory of Decentralized Resource Allocation," unpublished Ph.D. dissertation, Graduate School of Business, Stanford University.

Keren, M. [1972], "On the Tautness of Plans," Review of Economic Studies, October.

Kihlstrom, R., and M. Pauly [1971], "The Role of Insurance in the Allocation of Risk," American Economic Review, May.

Kleindorfer, P., and M. Sertel [1976a], "Optimal Design of Enterprises through Incentives," Preprint Series of the International Institute of Management, Berlin.

Kleindorfer, P., and M. Sertel [1976b], "Optimal Design of Labor-Managed Enterprises," Preprint Series, International Institute of Management, Berlin.

Kobayashi, T. [1977], notes, Graduate School of Business, Stanford University.

Laffont, J. J. [1977], "More on Prices vs. Quantities," Review of Economic Studies, February.

Loeb, M. [1975], "Coordination and Informational Incentive Problems in the Multidivisional Firm," unpublished Ph.D. dissertation, Northwestern University.

Luenberger, D. [1968], Optimization by Vector Space Methods, John Wiley & Sons, New York.

Marschak, J., and R. Radner [1972], Economic Theory of Teams, Yale: Yale University Press.

Mirrlees, J. [1971], "An Exploration in the Theory of Optimum Income Taxation," Review of Economic Studies, 38.

Mirrlees, J. [1974], "Notes on Welfare Economics, Information and Uncertainty," in Balch, McFadden, Wn, Essays on Economic Behavior under Uncertainty: North Holland, Amsterdam-London.

Mirrlees, J. [1976], "The Optimal Structure of Incentives and Authority within an Organization," Bell Journal of Economics, Spring.

Moulin, H. [1976], "Cooperation in Mixed Equilibrium," Mathematics of Operations Research, August.

Munkres, J. [1975], Topology: A First Course: Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Noreen, E. [1976], "Executive Stock Options: An Economic and Accounting Analysis," unpublished Ph.D. dissertation, Stanford University.

Pauly, M. [1974], "Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection," Quarterly Journal of Economics, February.

Prescott, E. [1972], "The Multi-Period Control Problem under Uncertainty," Econometrica, November.

Riley, J. [1976], "Informational Equilibrium," RAND Working Paper, December.

Ross, S. [1973], "The Economic Theory of Agency: The Principal's Problem," American Economic Review, May.

Rotschild, M., and J. Stiglitz [1976], "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," Quarterly Journal of Economics, November.

Samuelson, P. [1970], "The Fundamental Approximation Theorem of Portfolio Analysis in Terms of Means, Variances and Higher Moments," Review of Economic Studies, October.

Selten, R. [1974], "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," Working Paper No. 23, Institute of Mathematical Economics, University of Bielefeld, Germany.

Spence, M., and R. Zeckhauser [1971], "Insurance, Information and Individual Action," American Economic Review, May.

Spence, M. [1973], "Job Market Signalling," Quarterly Journal of Economics, August.

Spence, M. [1977], "Non-Linear Prices and Welfare," Journal of Public Economics, August.

Stiglitz, J. [1974], "Incentives and Risk Sharing in Sharecropping," Review of Economic Studies, April.

Stiglitz, J. [1975], "Incentives, Risk and Information: Notes toward a Theory of Hierarchy," Bell Journal of Economics, Autumn.

Stiglitz, J. [1976], "Information Economics," manuscript to a forthcoming book.

Thomson, W. [1976], "Incentives and Information," unpublished Ph.D. dissertation, Department of Economics, Stanford University.

Weitzman, M. [1974], "Prices vs. Quantities," Review of Economic Studies.

Weitzman, M. [1976a], "The New Soviet Incentive Model," The Bell Journal of Economics, Spring.

Weitzman, M. [1976b], "Optimal Revenue Functions for Economic Regulation," mimeo, M.I.T., October.

Wilson, R. [1968], "The Theory of Syndicates," Econometrica, 36.

Wilson, R. [1969], "The Structure of Incentives for Decentralization under Uncertainty," La Decision, No. 171.

Wilson, R. [1977], "Information, Efficiency and Core of an Economy" (forthcoming in Econometrica).

Yohe, G. [1977a], "Single-Valued Control over a Cartel under Uncertainty - A Multifirm Comparison of Prices and Quantities," Bell Journal of Economics, Spring.

Yohe, G. [1977b], "Single-Valued Control of an Intermediate Good under Uncertainty," International Economic Review, February.

Zeckhauser, R. [1970], "Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives," Journal of Economic Theory, 2.