MULTIFACETED ASPECTS OF AGENCY RELATIONSHIPS

by

ELLA MAE MATSUMURA

A.B., The University of California, Berkeley, 1974

M.Sc., The University of British Columbia, 1976

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Faculty of Commerce and Business Administration

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September 1984

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Commerce and Business Administration

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date September 25, 1984

ABSTRACT

Agency theory has been used to examine the problem of stewardship of an agent who makes decisions on behalf of a principal who cannot observe the agent's actual effort. Effort is assumed to be personally costly to expend. Therefore, if an agent acts in his or her own interests, there may be a "moral hazard" problem, in which the agent exerts less effort than agreed upon. This dissertation examines this agency problem when the agent's effort is multidimensional, such as when the agent controls several production processes or manages several divisions of a firm. The optimal compensation schemes derived suggest that the widely advocated salary-plus-commission scheme may not be optimal. Furthermore, the information from all tasks should generally be combined in a nonlinear fashion rather than used separately in compensating a manager of several divisions, even if the monetary outcomes are statistically independent.

In situations where effort is best interpreted as time, effort can be viewed as being additive. The analysis in this special case shows that the nature of the outcome distribution, including the effect of effort on the mean of the distribution, is critical in determining whether it is optimal for the principal to induce the agent to diversify effort across tasks. These new results and the already existing agency theory results are applied to the sales force management problem, in which the firm wishes to motivate a salesperson to optimally allocate time spent selling the firm's various products.

The agency model is also expanded to allow for the agent's observation of the first outcome (which is influenced by the agent's first effort) before choosing the second effort level. The optimal compensation schemes both in the absence of and the presence of a moral hazard problem are derived. The behavior of the second effort strategy is also examined. It is shown that the behavior of the agent's second effort strategy depends on the interaction between wealth and information effects of the first outcome. Results similar to those in the multidimensional effort case are obtained for the question of optimality of diversification of effort when effort is additive.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Professor Jerry Feltham amd Professor John Butterworth for the many hours they spent discussing ideas and results with me. Their comments often pointed me in new directions or helped me to look at problems from new perspectives. I am grateful to Professor Tracy Lewis for serving on my dissertation committee. I would also like to thank my husband, Kam-Wah Tsui, for his technical assistance while I was writing this dissertation, and for his unfailing encouragement and moral support during the entire time I was in the Ph.D. program. Finally, I grateful[y] acknowledge the financial support of the University of British Columbia and of the Social Sciences and Humanities Research Council of Canada.

CHAPTER 1

INTRODUCTION

Managerial accounting has traditionally been associated with the valuation of inventories for external reporting and with information provision for internal decision making and control. Broadly speaking, the internal decision making relates to the planning of operations and the control of decentralized organizations. Variance analysis, budgeting, cost-volume-profit analysis, and the development of performance evaluation measures are typical components of the planning and control processes.

There are a number of different approaches to gaining a better understanding of the role of the accounting system in the control of decentralized operations. Since an accounting system is an information system, any research on the value of information, the demand for information, or the roles or uses of information has potential implications for accounting research. The body of research which examines such information issues has come to be known as information economics. Information economics uses formal economic models in order to study the demand for information for decision making and performance evaluation purposes. In particular, information economics attempts to find economic explanations for why certain phenomena are observed (e.g., Demski and Feltham, 1978), and to uncover insights about behavior thought to be nonoptimal (e.g., Zimmerman, 1979) or behavior thought to be optimal (e.g., Baiman and Demski, 1980b).

Much of the early information economics literature focused on essentially single-person decision situations (e.g., Demski and Feltham, 1976, and Feltham, 1977a), where information serves only a decision-facilitating purpose. That is, the decision maker uses information about the uncertain state of nature to revise his or her beliefs about the decision environment. Thus, the demand for this type of information might be called <u>decision-mak-</u>

ing demand. The recent information economics literature has incorporated agency theory in explicitly modeling the multiperson nature of accounting problems (e.g., Baiman and Demski, 1980a, 1980b, Gjesdal, 1981, and Holmstrom, 1977). In multiperson situations, information can play a decision-influencing role. For example, if a manager's actions affect actual production costs, and the manager is evaluated and possibly compensated on the basis of the costs, then the manager's actions will be influenced by the existence of the information system which reports the costs. The demand for this type of information might be called performance evaluation demand, or stewardship demand.

This dissertation uses the agency framework to examine some of the issues in the development of performance evaluation measures for motivational purposes. The basic agency model provides a means of studying situations in which one individual (the principal) delegates the selection of actions to another individual (the agent). Within the context of the firm, the principal might be the employer or superior and the agent might be the employee or subordinate. The agency theory literature (e.g., Harris and Raviv, 1979, Holmstrom, 1979) uses the expected utility model to represent the preferences of the principal and the agent, and generally assumes that the agent's action (effort) and a random state of nature determine the monetary outcome. The sharing rule (contract or compensation scheme) offered by the principal to the agent specifies how much is paid to the agent for each possible value of some performance measure or measures. The performance measure is often taken to be the monetary outcome, or the monetary outcome and an imperfect signal about the agent's effort. The compensation can be based only on what is jointly observable to the principal and the agent, and the compensation must be adequate enough to induce the agent to work for the principal. Alternative employment opportunities for the agent are thus explicitly considered.

The principal will generally find it prohibitively costly to continuously monitor the agent to determine what action (effort) the agent chooses. Therefore, if the agent has disutility for effort and acts in his or her own self-interest, the potential for a moral hazard (incentive) problem exists because of the principal's inability to observe the agent's actions. If the principal pays the agent a fixed wage, the agent has no economic incentive to perform the agreed level of effort, since a low outcome can be blamed on a bad state of nature rather than on shirking by the agent. At the other extreme, if the principal rents capital or rents the firm to the agent for a fixed fee so that the agent gets the outcome less a fixed fee, the shirking problem can be avoided entirely. The shortcoming of this type of contract is that it imposes a nonoptimal amount of risk on the agent. That is, the principal and the agent could be made better off in an expected utility sense by using some other contract.

Agency theory provides a framework in which it is possible to find compensation schemes which efficiently motivate the agent to choose the desired actions. The idea is to create incentives through an employment contract which imposes <u>some</u> risk on the agent in order to provide incentives for the agent to expend some agreed level of effort. The consequences of the existence of <u>non</u>monetary returns or costs, such as effort, can thus be analyzed. This is important for the analysis of performance evaluation and managerial control systems, where incentive effects play a critical role. The choice of variables on which compensation is to be based can be formally derived, with implications for the design of information systems. Furthermore, the analysis clearly demonstrates how the information obtained can be incorporated for motivational purposes.

Most of the existing agency theory research (see Baiman (1982) for a comprehensive survey) has a rather narrow definition of effort, in that effort is assumed to be single-dimensional. However, people are often faced with several similar tasks which must be performed within one time period. Examples include a salesperson selling several products for a firm, an auditor allocating time to different tasks in an audit assignment, a manager controlling several production processes, or a manager overseeing several divisions of a company. The problem of motivating the optimal allocation of effort within one period is not only interesting in its own right, but also has possible implications for multiperiod problems, where effort is allocated across periods. Multiperiod problems are of interest because the eventual goal is to be able to analyze and understand the issues involved when there are current <u>and</u> long-term consequences of decisions, as there are in many accounting settings.

Chapter 2 of this dissertation contains the notation used in the remainder of the paper and a formulation of the agency problem with allocation of effort. Chapter 3 describes theoretical results and an application in the allocation setting, and Chapter 4 describes results in the one-period sequential choice setting. In this scenario, after each effort level is exerted, an associated outcome is observed by the agent before the next effort level is exerted. The agent is compensated only at the end of the sequence of outcomes. The one-period sequential choice case is an intermediate step between the allocation of effort case, in which both the efforts are exerted before the outcomes are known, and the multiperiod case, in which the first outcome is observed and the first compensation is paid before the second effort is exerted. Chapter 5 concludes the dissertation with an outline of proposed future research. All technical calculations and proofs appear in the appendices.

CHAPTER 2

NOTATION AND FORMULATION

In order to state the agency problem with allocation of effort, the following notation will be used:

R = the set of all real numbers,

R_+ = the set of all nonnegative real numbers,

X = the set of possible monetary outcomes,

$x \in X \subseteq R$ is the monetary outcome,

$\underline{x} = (x_1,\ldots,x_n)$ is a disaggregation of the monetary outcome x, i.e.,

$x = \sum_{i=1}^{n} x_i$, $w \in R^k$ is a k-dimensional vector-valued performance measure,

e.g., $w = \underline{x}$ with $k=n$,

$s(\cdot)$, a real-valued function, is a sharing rule over the arguments indicated, with $s(\cdot) \in [s_0,\bar{s}]$,[1]

a_i = effort expended on task i, $i=1,\ldots,d$,

$\underline{a} = (a_1,\ldots,a_d) \in A \subseteq R_+^d$,

$f(x,w|\underline{a})$ is the joint density of x and w conditional on $\underline{a}$, and is understood to be $f(\underline{x}|\underline{a})$ if w is a function of $\underline{x}$; $g(\cdot)$, $h(\cdot)$, and $\phi(\cdot)$ will also be used to denote probability distributions;

$U(\cdot)$: $R \to R$ is the agent's utility function over money, where $U' > 0$ and $U'' \leq 0$,

$V(\cdot)$: $R_+^d \to R$ is the agent's disutility function over effort, where $\partial V/\partial a_i > 0$ and $\partial^2 V/\partial a_i^2 > 0$,

$\bar{u}$ = the agent's minimum acceptable utility level,

$W(\cdot)$: $R \to R$ is the principal's utility function over money, where $W' > 0$ and $W'' \leq 0$,

argmax $\{\cdot\}$ = the set of arguments maximizing the expression in braces.

In order to avoid side-betting issues, it will be assumed that the principal and the agent have identical beliefs about the conditional probability distribution over the outcome and performance measure, given effort $\underline{a}$. As in much of the agency literature, the agent's utility function is assumed to be of the form $U(s) - V(\underline{a})$.[2] In most of the agency literature, $n=d=1$.

The principal's problem is

$$\underset{s(\cdot),\underline{a}}{\text{Maximize}} \iint W(x-s(w))f(x,w|\underline{a})\, dw\, dx \tag{2.1}$$

$$\text{subject to} \quad \iint [U(s(w))-V(\underline{a})]f(x,w|\underline{a})\, dw\, dx \geq \bar{u} \tag{2.2}$$

$$\underline{a} \in \operatorname{argmax} \left\{\iint [U(s(w))-V(\underline{a})]f(x,w|\underline{a})\, dw\, dx\right\}. \tag{2.3}$$

It will be assumed that (2.3) can be replaced with the conditions

$$\frac{\partial}{\partial a_i} \iint [U(s(w))-V(\underline{a})]f(x,w|\underline{a})\, dw\, dx = 0,\ i=1,\ldots,d. \tag{2.4}$$

Furthermore, sufficient regularity to allow differentiation inside the integral is assumed. This permits the replacement of (2.4) with

$$\iint U(s(w))f_{a_i}(x,w|\underline{a})\, dw\, dx = V_{a_i}(\underline{a}),\ i=1,\ldots,d, \tag{2.5}$$

with subscripts a_i denoting partial differentiation with respect to a_i.

The principal's problem is solved by means of a generalized Lagrangian technique. A Hamiltonian (Lagrangian) is formed by attaching a multiplier λ to (2.2) and multipliers μ_i to each constraint in (2.5). It will be assumed that the supports of x and w do not vary as $\underline{a}$ varies, and that the partial derivatives of f with respect to each a_i exist and are nondegenerate. The dimension d is often taken to be equal to n, and the marginal cumulative distribution functions are assumed to satisfy first order stochastic dominance. That is, if $F_i(x_i|a_i)$ is the marginal cumulative distribution function of x_i, then $\partial F_i(x_i|a_i)/\partial a_i \leq 0$, $i=1,\ldots,n$. In a framework where $x_i = h_i(a_i, \theta)$, where θ represents state uncertainty, if x_i is increasing in a_i (i.e., $\partial x_i | \partial a_i \geq 0$ for all θ), then $\partial F_i(x_i|a_i)/\partial a_i \leq 0$. Finally, the sharing rule is assumed to be measurable and bounded. For the most part, interior solutions will be examined.[3]

Some of the results will make use of two special classes of functions. The first is the HARA (hyperbolic absolute risk aversion) class of utility functions, whose risk aversion functions are of the form

$$- U''(x)/U'(x) = 1/(Cx + D). \tag{2.6}$$

The $C = 1$ case corresponds to $U(x) = \ln(x+D)$, the $C = 0$ case corresponds to $U(x) = - \exp[-x/D]$, and the other cases correspond to power utility functions.

The other class of interest is the one-parameter exponential family of distributions. This class includes the exponential, gamma (with the shape parameter fixed), normal (with constant variance), and Poisson distributions. The following representation differs slightly from the usual one for a one-parameter exponential family (see, e.g., DeGroot, 1970).

Definition: A probability density function $f(x|a)$ with respect to the measure $r(.)$ will be said to belong to the one-parameter exponential family Q if it can be written as

$$f(x|a) = \exp[z(a)x - B(z(a))]h(x), \tag{2.7}$$

where $r(.)$ is the Lebesgue measure when the random variable x is absolutely continuous, and $r(.)$ is some counting measure when x is discrete.

The representation in (2.7) has the advantage that closed-form expressions can be obtained for $E(x|a)$ and $Var(x|a)$. In particular, $E(x|a) = B'(z(a))$ and $Var(x|a) = B''(z(a))$ (Peng, 1975). Table II in Appendix 1 details the representations of some familiar distributions. The remainder of Appendix 1 consists of calculations which are useful in the proofs of the results in Chapters 3 and 4.

CHAPTER 2 FOOTNOTES

1. If the sharing rule is unbounded, an optimal solution may not exist (Mirrlees, 1974; Holmstrom, 1977, 1979). Furthermore, the agent's wealth places bounds on the possible sharing rules.

2. Gjesdal (1981) has shown that such a utility function for the agent ensures that nonrandomized payment schedules are Pareto optimal. His result refers to ex post (after effort selection by the agent) randomization only. Fellingham, Kwon, and Newman (1983) have shown that ex ante randomization of payment schedules is optimal under certain conditions. It will be assumed in what follows that these conditions are not satisfied, and hence the focus is on pure (nonrandomized) payment schedules.

3. That is, the focus will be on the first-order conditions, which apply to interior solutions.

CHAPTER 3

ALLOCATION OF EFFORT

As stated earlier, the agency theory framework explicitly recognizes alternative employment opportunities for the agent, disutility for effort, risk aversion of the agent, and the possibility of the principal obtaining information about the agent's effort, all for situations in which the agent has one task to perform. However, in many situations, job effort is multidimensional; the agent must allocate effort to several different, but possibly related tasks. In spite of the variety of situations in which multidimensional job effort occurs, little attention has been devoted to characterizing optimal compensation schemes for these situations. Stiglitz (1975) considered multidimensional job effort under linear incentive schemes, and Weinberg (1975) sought an incentive compatible scheme for the problem of sales force management in multiproduct firms. Radner and Rothschild (1975) examined the properties of three heuristic strategies an agent might employ when faced with the problem of allocating effort. More recently, Gjesdal (1982) allowed for multidimensional effort and focused on the value of information.

The focus of this chapter is the characterization of optimal incentive schemes for the agency problem with allocation of effort across several tasks. The issues of separability of the optimal sharing rule across tasks and the value of additional information are examined, and the results suggest that certain compensation schemes that are widely advocated may not be optimal. In particular, commission schemes and linear sharing rules are shown not to be optimal, in general. The special case of additive effort is discussed, and the results are applied to the problem of sales force management.

3.1 FIRST BEST

Suppose that in addition to observing the aggregated or disaggregated outcome (i.e., $w = x$ or $w = \underline{x}$), the principal can observe the agent's effort. These cases may be called complete contractual information cases, since the principal can observe the agent's choice of effort. These "first best" situations are interesting as benchmarks for comparison with "second best" situations, those in which there is less than complete contractual information. The characterizations of the optimal sharing rule for these first best cases are obtained by solving the problem given by (2.1) and (2.2).

As in the single-dimensional effort case, if one individual is risk neutral and the other is risk averse, then the risk neutral individual bears all the risk. Thus, if the agent is risk neutral ($U(s) = s$) and the principal is risk averse, then Pareto optimal sharing rules are $s(x) = x - k$ and $s(\underline{x}) = x - k$, where k is a fixed fee paid to the principal. Conversely, if the principal is risk neutral and the agent is not, the principal bears the risk, receiving a share of $x - c$, while the agent receives a constant wage c. In the event that both the principal and the agent are risk neutral, the a_i's are chosen so that the agent's marginal disutility for effort equals the marginal increases in the expected outcome (i.e., so that $\partial E(x|a)/\partial a_i = \partial V/\partial a_i$, $i=1,\ldots,d$), and the sharing rule can be taken as $s(\cdot) = \bar{u} + V(a^*)$, with the principal receiving $E(x|a^*) - \bar{u} - V(a^*)$.

If both the agent and the principal are risk averse, then they each bear part of the risk, as indicated in Proposition 3.1.1 below.

Proposition 3.1.1. If both the agent and the principal are risk averse and they have homogeneous beliefs, then $s(\underline{x})$ varies only with x in the first best case.

Because the optimal sharing rule depends only on x, $s(\underline{x})$ is the same for all $\underline{x}$ that provide the same total x. The sharing rule $s(\underline{x})$ therefore varies with x only for risk-sharing purposes - the makeup of x is unimportant. Moreover, it is easily seen that $s(\underline{x})$ is increasing in each x_i, regardless of the properties of the conditional distribution function on $\underline{x}$. This is in contrast to the second best solution.

3.2 SECOND BEST

Suppose now that the principal cannot observe the agent's effort, and hence must present the agent with a sharing rule which induces the desired choice of effort. Since the focus in most of what follows is on motivational, rather than risk-sharing issues, it will be assumed that unless otherwise stated the principal is risk neutral and the agent is risk averse. As remarked above, if there were no moral hazard problem, the principal would then bear all the risk. Whatever risk is imposed on the agent in the second best case is thus imposed not for risk-sharing purposes, but rather for motivational purposes.

Letting f_{a_i} denote $\partial f/\partial a_i$, the optimal sharing rule, given that only x is observed, is characterized by

$$\frac{1}{U'(s(x))} = \lambda + \frac{\sum_{i=1}^{d} \mu_i f_{a_i}(x|\underline{a}^*)}{f(x|\underline{a}^*)}, \tag{3.2.1}$$

for almost every x such that $s(x) \in [s_0, \bar{s}]$. For all other x, $s(x) = s_0$ if the left hand side of (3.2.1) is greater than the right hand side, and $s(x) = \bar{s}$ if the opposite is true.

For example, suppose that n=2, $U(s) = \ln s$, x_1 and x_2 are independent, $x_1 \sim N(a_1, \sigma_1^2)$, and $x_2 \sim N(a_2, \sigma_2^2)$. Then $x \sim \underset{\sim}{N}(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$ and $1/U'(s) = s$. The interior portion of the optimal sharing rule is thus[1] (See Table II in Appendix 1 for the normal density)

$$s(x) = \lambda + (\mu_1 + \mu_2)\frac{x - a_1^* - a_2^*}{\sigma_1^2 + \sigma_2^2}$$

$$= \lambda - \frac{(\mu_1 + \mu_2)(a_1^* + a_2^*)}{\sigma_1^2 + \sigma_2^2} + \frac{\mu_1 + \mu_2}{\sigma_1^2 + \sigma_2^2}x,$$

which can be interpreted as a compensation scheme consisting of a fixed portion plus a commission. If the agent's utility function is $U(s) = 1-e^{-s}$, then the interior portion of the optimal sharing rule is

$$s(x) = \ln[\lambda + (\mu_1 + \mu_2)\frac{x - a_1^* - a_2^*}{\sigma_1^2 + \sigma_2^2}].$$

In general, if the principal is risk neutral and the agent's utility function is in the HARA class, with risk aversion function given by $-U''(s)/U'(s) = 1/(Cs+D)$, then the interior portion of the optimal sharing rule is

$$s(x) = \begin{cases} \frac{1}{C}[(\lambda + \frac{\sum_{i=1}^{d} \mu_i f_{a_i}(x|\underline{a})}{f(x|\underline{a})})^C - D], & \text{if } C \neq 0 \\ D\ln(\lambda + \frac{\sum_{i=1}^{d} \mu_i f_{a_i}(x|\underline{a})}{f(x|\underline{a})}), & \text{if } C = 0. \end{cases} \qquad (3.2.2)$$

($C=1$ corresponds to $U(s) = \ln(s+D)$, $C=0$ corresponds to $U(s) = -e^{-s/D}$, and the other cases correspond to power utility functions.)

As in the single-task setting, the first best solution is achievable with a fixed fee going to the principal when the agent is risk neutral. This can be deduced by interpreting effort to be a vector rather than a scalar in the single-task setting proofs (e.g., Shavell, 1979). Thus, even though less than complete contractual information is available, the principal and the agent can obtain the same expected utilities as they could in the complete contractual information case. This is because in effect, the risk-neutral agent rents the firm from the principal for a fixed fee.

When U(.) is strictly concave, equation (3.2.1) implies that a necessary and sufficient condition for s(x) to be nondecreasing in x is[2]

$$\sum_{i=1}^{d} \mu_i \frac{\partial}{\partial x}\left[\frac{f_{a_i}(x|\underline{a}^*)}{f(x|\underline{a}^*)} \right] \geq 0 \tag{3.2.3}$$

for all x corresponding to interior solutions, where the μ_i's are the Lagrangian multipliers associated with the optimal solution (a^*,s(x)). When a is one-dimensional, (3.2.3) reduces to

$$\frac{\partial}{\partial x}\left[\frac{f_a(x|a^*)}{f(x|a^*)} \right] \geq 0, \tag{3.2.4}$$

since μ is positive (Holmstrom, 1979). If (3.2.4) is true for all $a^* \in A$, then f(x|a) has the monotone likelihood ratio property in x (Lehmann, 1959, p. 111). Many distributions, including all those in the one-parameter exponential family, have the monotone likelihood ratio property; this property is a stronger ordering on distributions than is first-order stochastic dominance (Lehmann, 1959, pp. 73-74).[3] If $\underline{a}$ is multidimensional and the Lagrangian multipliers μ_i are all nonnegative, then a sufficient condition for s(x) to be nondecreasing in x is

$$\frac{\partial}{\partial x}\left[\frac{f_{a_i}(x|\underline{a}^*)}{f(x|\underline{a}^*)} \right] \geq 0, \text{ for all x, i=1,...,d.}$$

When a is single-dimensional, the first-order stochastic dominance property means that as a increases, the distribution f(x|a) shifts to the right. It is this property that accounts for the monotonicity of the optimal sharing rule. When $\underline{a}$ is multidimensional, the problem of determining an ordering over the effort vectors arises. Condition (3.2.3) states that the directional derivative of log $f(x|\underline{a})$ in the direction of $\mu = (\mu_1,\ldots,\mu_d)$ at

$\underline{a}^*$ be nonnegative in order for the optimal sharing rule to be monotonic. Thus, μ provides a direction in which to measure the shifting of $f(x|\underline{a})$.

Because of the critical role that the multipliers μ_i play in the sharing rule, it is of interest to try to determine whether they are strictly positive. A partial answer is provided in Proposition 3.2.1 below, for the situation in which the vector $\underline{x} = (x_1,x_2)$ is observed.

Proposition 3.2.1. Suppose $\underline{x} = (x_1,x_2)$, $f(\underline{x}|\underline{a}) = g(x_1|a_1)h(x_2|a_2)$, and $F_{a_i}(x_i|a_i) \leq 0$, $i=1,2$, with strict inequality for some x_i values. Suppose further that the agent's expected utility is strictly concave in $\underline{a}$. Then at least one of μ_1 and μ_2 must be positive.

It can be shown that if the agent's utility for wealth is $U(s) = 2\sqrt{s}$, $V(\underline{a})$ is strictly convex in $\underline{a}$, and $g(\cdot)$ and $h(\cdot)$ are exponential distributions with means a_1 and a_2, respectively, then the agent's expected utility is strictly concave in $\underline{a}$. In this case, if the principal is risk neutral, then both μ_1 and μ_2 are positive. Proposition 3.5.9 in section 3.5 provides other conditions under which both μ_1 and μ_2 are positive.

A final remark on the characteristics of the optimal sharing rule can be made at this point. Suppose $f(\underline{x}|\underline{a})$ is of the form given in Proposition 3.2.1, i.e., $f(\underline{x}|\underline{a}) = \prod_{i=1}^{d} g^i(x_i|a_i)$, where the superscripts on $g(\cdot)$ are merely indices. The optimal sharing rules is characterized as in (3.2.1), with x replaced by $\underline{x}$. In the special case under consideration, this reduces to $\frac{1}{U'(s(\underline{x}))} = \lambda + \sum_{i=1}^{d} \mu_i g_{a_i}(x_i|a_i^*)/g(x_i|a_i^*)$. If, further, each $g^i(\cdot)$ belongs to the one-parameter exponential family Q described by (2.7), then each $g^i_{a_i}(\cdot)/g^i(\cdot)$ is a constant multiplied by $(x_i - M_i(a_i))$, where $M_i(a_i)$ is the mean of x_i given a_i. The means $M_i(a_i)$ can be thought of as standards or norms, so that the optimal sharing rule is a function of deviations from standards (cf. Christensen, 1982). This is consistent with managerial accounting's focus on variances (deviations from standards) as an aid in performance evaluation.

3.3 VALUE OF ADDITIONAL INFORMATION

A question which naturally arises at this point is: Would the principal be better off knowing each x_i rather than only x? That is, will the principal always be strictly better off with disaggregated or finer information? More generally, under what conditions will the principal or the agent be made strictly better off by information in addition to the aggregate outcome, x?

Intuitively, the more (imperfect) information the principal has about the agent's effort, the more efficiently the agent can be motivated to exert effort. Consequently, the principal's expected utility should increase in most situations where additional information is available. A number of people have addressed this problem. Holmstrom (1979), for example, showed that if the additional information is of _value_ (that is, if its optimal use will lead to a Pareto superior pair of expected utilities for the principal and the agent), then the additional information must be _informative_ in the sense that it contains information about the agent's effort that is not contained in the output. The converse was also shown to be true. Gjesdal (1981, 1982) examined the relationship between Blackwell (1953) informativeness and the value of information.

In order to define Blackwell informativeness, let Ω be a set of possible performance measures ω. In this section, ω is assumed to include the output, x. An information system η is a function from Ω to some signal space Y. Let y denote an arbitrary element in Y and let A, the set of all possible actions, be finite. Information system η is _Blackwell more informative_ than another system $\gamma : \Omega \to Z$ if and only if $P_\gamma(z|a) =$

$\int_Y P(z|y)dP_\eta(y|a)$ for each action a in A and each signal z in Z. It should be noted that although η is said to be Blackwell more informative than γ, η may actually be only equally informative as γ is. Amershi (1982, Appendix 1) has generalized the definition for the case where A is infinite.

Amershi (1982) re-examined the value of additional information problem, and corrected and generalized the results of Holmstrom (1979, 1982) and Gjesdal (1981, 1982). Amershi (1982) showed that a risk neutral principal weakly prefers an information system that is Blackwell more informative than another. That is, the principal's and the agent's expected utilities are at least as high with the Blackwell more informative system than with the other. A risk averse principal requires that the Blackwell more informative system also provide a specific form of information about the output (see Proposition 3.3.1 below). These results differ from the single-person decision maker case, where risk attitudes are immaterial. Intuitively, a risk neutral principal is concerned only with the incentive properties of contracts, whereas a risk averse principal is concerned with both the incentive and risk-sharing properties of contracts. The risk-sharing aspect accounts for the conditions on the output in Proposition 3.3.1 below.

More specifically, Proposition 3.3.1 says that information system η is at least as preferred as information system γ if η is Blackwell more informative than γ with respect to the effort, a, and (i) there is no risk-sharing involved, or (ii) γ says nothing more about the output x than η does, or (iii) the signal provided by η is enough to determine the output.

Proposition 3.3.1 (Amershi (1982, Theorem 3.1)). Let an information system $\eta : \Omega \to Y$ be more informative in the Blackwell sense than the system $\gamma : \Omega \to Z$ with respect to the family of measures $P_A = \{P(\omega|a) : a \in A\}$. Suppose also, at least one of the following conditions hold: (i) The principal is risk neutral. (ii) The output variable and the information system γ are conditionally independent given η. (iii) The output can be expressed as $x = h(\eta(\omega))$ for some measurable function $h : Y \to R$. Then the principal weakly prefers η over γ.

In this proposition and in the other propositions in this section, ω is a vector of performance measures that includes the output, x. Although the effort variable, a, is taken to be single dimensional, the proof holds for finite-dimensional effort <u>vectors</u> as well.

Proposition 3.3.1 identifies conditions under which information system η is at least as preferred to information system γ. It is of interest to identify conditions under which η is strictly preferred to γ.

Amershi's (1982) strict preference results rely on the concept of sufficient statistics. Using the notation above, a statistic $T : \Omega \to K$ is <u>sufficient</u> for the family of measures $P_A = \{P(\omega|a) : a \varepsilon A\}$ if and only if there exists a nonnegative function $h : \Omega \to R_+$ and functions $g(\cdot|a) : K \to R$ such that

$$f(\omega|a) = h(\omega)g(T(\omega)|a) \text{ for all } \omega\varepsilon\Omega \text{ and } a\varepsilon A,$$

where $f(\cdot)$ is a density if the random variable is continuous, or a mass function if the random variable is discrete. A sufficient statistic may be viewed as an information system. A <u>minimal sufficient statistic</u> is a sufficient statistic $\hat{T} : \Omega \to L$ that is a function of every other sufficient statistic. An <u>agency sufficient statistic</u> (Amershi, 1982) Ψ is equal to a sufficient statistic T on Ω if the principal is risk neutral, or (X,T) if the principal is risk averse. Ψ is called a <u>minimal agency sufficient statistic</u> if the sufficient statistic T is minimal. Finally, a contract (s^*,a^*) is called a <u>best agency contract</u> if there is no other contract based on any information system on Ω that is strictly preferred to it.

The proposition below uses the concept of agency sufficient statistics to characterize strict preferences for information systems. Essentially, the principal will strictly prefer an agency sufficient statistic η over another system γ which does not generate a best contract. This is because a best contract must be a function of the minimal agency sufficient statistic, which extracts all relevant information from ω about a (Amershi (1982, Corollary 3.3)). Proposition 3.3.2 below provides a situation in which the information system γ cannot generate a best contract.

Proposition 3.3.2 (Amershi (1982, Proposition 3.4)). Suppose a best contract exists and at each best agency contract,

$$\frac{W'(x-s^*_\beta(\beta(\omega)))}{U'(s^*_\beta(\beta(\omega))} = \lambda + \mu \frac{\partial}{\partial a} \log f(\omega|a^*_\beta),$$

where β is an information system which leads to a best agency contract. The principal strictly prefers an agency sufficient statistic η over a system γ if $\frac{\partial}{\partial a} \log f(\omega|a^*_\gamma)$ is not a function of γ if the principal is risk neutral (or not a function of (x, γ) if the principal is risk averse). Here (s^*_γ, a^*_γ) is the optimal contract based on γ.

Proposition 3.3.2 holds for the multidimensional effort case, with $\mu \frac{\partial}{\partial a} \log f(\omega|a^*_\gamma)$ replaced by $\sum_{i=1}^{n} \mu_i \frac{\partial}{\partial a_i} \log f(\omega|\underline{a}^*_\gamma)$. The following corollaries of Proposition 3.3.2 for the multidimensional effort case are modifications of Amershi's (1982) corollaries to his Proposition 3.4. Their proofs are immediate. Corollary 3.3.3 deals with the situation in which an additional signal z would be of positive value given an information system which reports the outcome x and another signal y. As in Proposition 3.3.2, a condition is provided which implies that $\gamma(x,y,z) = (x,y)$ cannot generate a best contract. Since $\eta(x,y,z) = (x,y,z)$ is trivially a sufficient statistic, Corollary 3.3.3 follows directly from Proposition 3.3.2. Corollary 3.3.4 provides a situation in which a sufficient statistic is strictly preferred to a nonsufficient statistic.

Corollary 3.3.3 (Gjesdal (1982, Proposition 1)). Let $\Omega = \{\omega = (x,y,z) :$ x,y,z are from some spaces$\}$. Let η be the information system that reports (x,y,z), and let γ be the information system that reports (x,y). Assume that for $\beta = \eta$ and $\beta = \gamma$,

$$\frac{W'(x-s^*_\beta(\beta(\omega)))}{U'(s^*_\beta(\beta(\omega)))} = \lambda + \sum_{i=1}^{n} \mu_i \frac{\partial}{\partial a_i} \log f(\omega|\underline{a}^*_\beta) \tag{3.3.1}$$

holds at the contracts $(\eta, s^*_\eta, a^*_\eta)$ and $(\gamma, s^*_\gamma, a^*_\gamma)$. Then the signal z has marginal value given (x,y) (that is, the principal strictly prefers η over γ) if

$$\sum_{i=1}^{n} \mu_i \frac{\partial}{\partial a_i} \log f(\omega|\underline{a}^*_\gamma) \tag{3.3.2}$$

is not a function of (x,y).

Corollary 3.3.4 (Holmstrom (1982, Theorem 6)). Suppose the principal is risk neutral, and suppose that for some system $\gamma : \Omega \to Z$, the expression in (3.3.2) is not a function of γ at each $\underline{a} \in A$. Then the principal strictly prefers any sufficient statistic η over γ if equation (3.3.1) holds at any best agency contract generated by information system β.

As Amershi (1982) remarks, these strict preference results do not establish that an agency sufficient statistic is always strictly preferred to a nonsufficient statistic. In order for a sufficient statistic to be strictly preferred to a nonsufficient statistic, the principal must use information which is provided by the sufficient statistic but not provided by the nonsufficient statistic. In addition, the principal's risk attitude is a factor, as shown in the proposition below. Part (2) of Proposition 3.3.5 says that sufficiency alone cannot determine strict preference ordering of information systems if the principal is risk averse.

Proposition 3.3.5 (Amershi (1982, Proposition 3.5)). Let η be the minimal sufficient statistic and x be the output.

(1) A risk neutral principal strictly prefers η over any system γ which is not a sufficient statistic if and only if every best agency contract (s^*,a^*) is such that s^* is a sufficient statistic.

(2) A risk averse principal strictly prefers (x, η) over any system γ such that (x, γ) is not an agency sufficient statistic if and only if every best agency contract (s^*,a^*) is such that (x,s^*) is an agency sufficient statistic.

Again, although the effort variable is single-dimensional, the result holds even if effort is multidimensional.

Amershi (1982) next developed the following result. Suppose $\eta(\omega)$ is a minimal sufficient statistic. If $\frac{\partial}{\partial a} \log f(\omega|a^*) \equiv \frac{\partial}{\partial a} \log k(\eta(\omega)|a^*)$ is an invertible function of $\eta(\omega)$, then a risk neutral principal strictly prefers η over any system γ that is not a sufficient statistic, and a risk averse principal strictly prefers (x, η) over any system (x, γ) which is not an agency sufficient statistic.

Unlike Amershi's (1982) previous results, which were easily extended to the multidimensional effort case, the invertibility result above does not lend itself to the multidimensional effort case. Intuitively, the dimension of a sufficient statistic cannot be less than the dimension of the vector of parameters to be estimated. For example, suppose that $x_1,\ldots,x_n$ $(n \geq 2)$ are observations from a normal distribution with unknown mean θ and unknown variance σ^2. Then a sufficient statistic for the vector of parameters (θ,σ^2) is $(\bar{x},s^2)$, where $\bar{x}$ is the sample mean and s^2 is the sample variance. Moreover, it is obvious that more than one observation is needed in order to make inferences about (θ,σ^2). Thus, a sufficient statistic in the multidimensional effort case will generally be multidimensional. The impossibility of inverting a one-dimensional value to obtain a multidimensional statistic precludes the use of Amershi's invertibility result in the allocation of effort problem.

For example, in the allocation of effort problem, $\underline{x} = (x_1,\ldots,x_n)$ is potentially observable, with the distribution of $\underline{x}$ parameterized by $\underline{a} = (a_1,\ldots,a_d)$. The statistic $x = \sum_{i=1}^{n} x_i$ can only be sufficient for $(x,\underline{x})$ if $\underline{a}$ is not really multidimensional, i.e., if there is some known functional relationship among the a_i's so that knowledge of one a_i is sufficient to perfectly infer the others. A special case of this type of relationship occurs when it is known that the agent will always choose the a_i's to be equal. In the allocation problem, it is very unlikely that $\underline{a}$ is not really multidimensional, and therefore in general, x is <u>not</u> sufficient for $(x,\underline{x})$, i.e., the minimal sufficient statistic is multidimensional.

Continuing with the focus on the value of additional disaggregated information, the principal's weak preference for the additional information is easily established. A multidimensional-effort version of Proposition 3.3.1 shows that the information system reporting $\underline{x} = (x_1,\ldots,x_n)$ is weakly preferred to the information system reporting only $\sum_{i=1}^{n} x_i$, no matter what the principal's or the agent's risk attitudes are.

If the principal can observe $\underline{x}$, the interior portion of the optimal sharing rule is characterized by

$$\frac{1}{U'(s(\underline{x}))} = \lambda + \sum_{i=1}^{d} \mu_i \frac{g_{a_i}(\underline{x}|\underline{a})}{g(\underline{x}|\underline{a})}.$$

To illustrate, suppose again that n=2, and U(s) = ln s, but let $\underline{x} = (x_1,x_2) \sim N(\underline{a},\Sigma)$, where Σ is the covariance matrix $\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$.

Then the interior portion of the optimal sharing rule is (see Appendix 2 for bivariate normal calculations)

$$s(\underline{x}) = \lambda + \sum_{i=1}^{2} \mu_i \frac{g_{a_i}(\underline{x}|\underline{a})}{g(\underline{x}|\underline{a})},$$

$$= \lambda + \mu_1\left[\frac{x_1-a_1}{\sigma_1^2(1-\rho^2)} - \frac{\rho(x_2-a_2)}{\sigma_1\sigma_2(1-\rho^2)}\right] + \mu_2\left[\frac{x_2-a_2}{\sigma_2^2(1-\rho^2)} - \frac{\rho(x_1-a_1)}{\sigma_1\sigma_2(1-\rho^2)}\right],$$

$$= \lambda + (x_1-a_1)\left[\frac{\mu_1}{\sigma_1^2(1-\rho^2)} - \frac{\rho\mu_2}{\sigma_1\sigma_2(1-\rho^2)}\right] + (x_2-a_2)\left[\frac{\mu_2}{\sigma_2^2(1-\rho^2)} - \frac{\rho\mu_1}{\sigma_1\sigma_2(1-\rho^2)}\right].$$

This compensation scheme may be interpreted as a commission scheme with different commission rates for each task. If $\sigma_1 = \sigma_2$ and $\mu_1 = \mu_2$, the commission rates will be the same for both tasks. It should be noted that in general, even if x_1 and x_2 are independent, the optimal commission rates need not be equal across tasks. This is because when x_1 and x_2 are independent (ρ=0),

$$s(\underline{x}) = \lambda - a_1\mu_1/\sigma_1^2 - a_2\mu_2/\sigma_2^2 + x_1\mu_1/\sigma_1^2 + x_2\mu_2/\sigma_2^2.$$

In this case, the commission rate for task i depends only on the variance of x_i and the multiplier μ_i. Since the sharing rule depends on each x_i, the signal $\underline{x}$, obtained in addition to x, is valuable (unless $\mu_1/\sigma_1^2 = \mu_2/\sigma_2^2$). This can be deduced formally from Proposition 3.3.2.

3.4 ADDITIVE SEPARABILITY OF THE SHARING RULE

Once the possibility of observing each x_i is introduced, the question of whether or not to reward the agent for each outcome separately arises. For example, should a manager of two divisions that are geographically dispersed be rewarded for the performance of each separately? Analytically, the question is whether the optimal sharing rule is additively separable in the x_i's. This question will be addressed for the HARA class of utility functions.

Let $\nabla(\underline{x}) = \lambda + \Sigma\, \mu_i g_{a_i}(\underline{x}|\underline{a})/g(\underline{x}|\underline{a})$. As before, if the agent's utility function is in the HARA class, with $-U''(s)/U'(s) = 1/(Cs+D)$, then the interior portion of the optimal sharing rule is given by

$$s(\underline{x}) = \begin{cases} \frac{1}{C}((\nabla(\underline{x}))^C - D), & \text{if } C \neq 0 \\ D \ln(\nabla(\underline{x})), & \text{if } C = 0, \end{cases} \tag{3.4.1}$$

for almost every $\underline{x}$ such that $s(x) \in [s_0, \bar{s}]$. If the principal is risk averse, with utility function in the HARA class and with identical cautiousness C (see (2.6)), then the interior portion of the optimal sharing rule is

$$s(\underline{x}) = \begin{cases} \dfrac{(\nabla(\underline{x}))^C (Cx+D_1) - D_2}{C(1 + (\nabla(\underline{x}))^C)}, & \text{if } C \neq 0 \\ \dfrac{D_1 D_2 \ln \nabla(\underline{x}) + D_2 x}{D_1 + D_2}, & \text{if } C = 0, \end{cases} \tag{3.4.2}$$

where D_1 corresponds to the principal, and D_2 corresponds to the agent.

Equation (3.4.1) implies that if the principal is risk neutral and the agent's utility function is in the HARA class, then a necessary condition for the optimal sharing rule to be additively separable is that C=1, i.e., that the agent have a log utility function. Given that $U(s) = \ln s$, a strong form of independence of the outcomes, $x_1,\ldots,x_n$, is a sufficient condition for the optimal sharing rule to be additively separable. More specifically, let $g^i(x_i|a_i)$ be the density of outcome x_i given effort a_i, and let $g(\underline{x}|\underline{a})$ be the joint density of $\underline{x}$ given $\underline{a}$. Then

$$g(x_1,\ldots,x_n|a_1,\ldots,a_n) = \prod_{i=1}^{n} g^i(x_i|a_i) \tag{3.4.3}$$

is a sufficient condition for additive separability of the optimal sharing rule, given that $U(s) = \ln s$. In this case, $s(\underline{x}) = \lambda + \sum_{i=1}^{n} s^i(x_i)$, where $s^i(x_i) = \mu_i \frac{g^i_{a_i}(x_i|a_i)}{g^i(x_i|a_i)}$. The example in Section 3.3 shows that given $U(s) = \ln s$, independence is a sufficient but not a necessary condition for separability of the sharing rule. One might conjecture that there are other common distributions of dependent random variables which, when $U(s) = \ln s$, yield a separable sharing rule. However, no other common joint distributions which seem appropriate (see, e.g., Johnson and Kotz, 1972), seem to lead to such a result. In general, then, the optimal sharing rule will <u>not</u> be additively separable.

It is interesting to note that (3.4.3) is not sufficient to yield a separable sharing rule if $U(s) \neq \ln s$. Furthermore, (3.4.3) is not sufficient to yield a separable sharing rule if both the principal and the agent are risk averse, with HARA-class utility functions and identical cautiousness C. This is easily seen from equations (3.4.2). Hence, even if the principal and agent have identical log utility functions, a separable sharing rule is not optimal.

These results differ from those in the cooperative setting, in which a weighted sum of the principal's and the agent's expected utilities is maximized (no Nash constraint is necessary). In the cooperative case, if beliefs are identical and the principal and agent are strictly risk averse, then the optimal sharing rule is linear for all weights if and only if the individuals have HARA-class utilities with identical cautiousness (Amershi and Butterworth, 1981). Thus, the moral hazard problem partially accounts for the generally nonlinear form of the optimal sharing rules.

One additively separable compensation scheme which is commonly used is the commission scheme. This scheme has the further restriction that

$s^i(x_i) = c_i x_i + b_i$, a linear function of x_i. As above, a necessary condition for a commission scheme (linear sharing rule) to be optimal is that the principal be risk neutral and that the agent have a log utility function. Given $U(s) = \ln s$, whether or not the optimal sharing rule is linear depends on the conditional distribution of the outcomes given effort.

3.5 ADDITIVE EFFORT

This section examines the special case where effort is additive, as when effort represents time spent on different tasks, and where there is no intrinsic disutility for any particular task. In this case, the disutility function for effort expended on d tasks can be written as $V(a_1+\ldots+a_d)$. This case necessitates only minor changes in the analysis; partial derivatives of $V(.)$ with respect to a_i are replaced by $V'(.)$. The assumption that the principal is risk neutral and the agent is risk averse will be maintained in this section.

Suppose that there is one outcome x_i associated with each a_i, and that the mean of each x_i is $k_i m_i(a_i)$, so that $E(x) = \sum_{i=1}^{d} k_i m_i(a_i)$, where $k_i > 0$ and $m_i'(\cdot) > 0$. In the first best case, the first order conditions require that the agent receive a constant wage and that

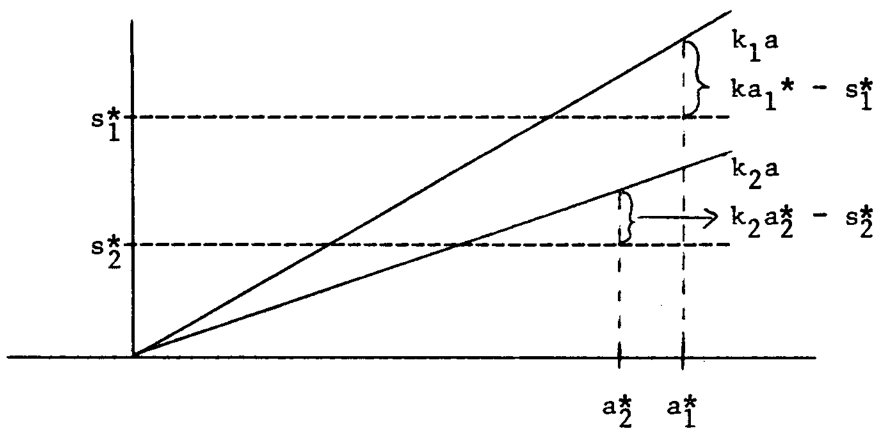
$$\partial E(x|\underline{a})/\partial a_i \equiv k_i m_i'(a_i) = \lambda V'(.), \text{ for all } i. \qquad (3.5.1)$$

The simplest case is that of constant marginal productivity, where $m_i(a_i) = a_i$, for all i. If, further, $k_i = k$ for all i, then (3.5.1) indicates that

$$k/\lambda = V'(\Sigma a_i), \qquad (3.5.2)$$

and hence any mix of efforts satisfying (3.5.2) is equally acceptable to both the principal and the agent. The k_i's may be thought of as measures of efficiency of effort (Shavell, 1979). If all the k_i's are unequal, then a boundary solution results. In particular, all but one of the a_i's are zero. The problem is thus essentially one of choosing on which task of many to expend effort. Suppose there are two tasks, with $k_1 > k_2$. In this situation, the optimal solution is to devote effort exclusively to task one. These results are summarized as Lemma 3A.2 in Appendix 3, where the proofs can also be found.

Comparison of two one-dimensional effort situations with $k_1 > k_2$ shows why the principal is better off with $a_1^* > 0$ and $a_2^* = 0$ than with $a_1^* = 0$ and $A_2^* > 0$. Since $k_1 > k_2$, there is a higher return per unit of effort for task one than from task two. Furthermore, it is worthwhile for the principal to induce more effort for task one than for task two (see Proposition 3A.3 and its proof in Appendix 3). The combined productivity gains (recall that $E(x_i) = k_i a_i$) outweigh the required increased fixed wage compensation to the agent, who would receive the same expected utility for either task. The principal's situation can be depicted graphically as follows:



For general $m_i(a_i)$, (3.5.1) implies that $k_i m_i'(a_i) = k_j m_j'(a_j)$, $i,j=1,\ldots,d$. The marginal impacts of the a_i's on the expected outcomes are balanced, and hence the solution will generally be interior. If the mean functions are identical, then the optimal efforts will be equal.

Although the agent's utility for wealth is not important in determining the principal's choice of the a_i's in the first best case, it is important in the second best case. Assuming an interior solution, the first order conditions in the second best case require that

$$\frac{\partial EU(s(x))}{\partial a_i} = \frac{\partial EU(s(x))}{\partial a_j}, \qquad i,j=1,\ldots,d \tag{3.5.3}$$

Since the agent's effort is not observable in this case, the principal must induce the agent to exert the optimal amount of effort at one or more tasks. The principal may find it optimal to devote resources to preventing shirking at only one task even if multiple tasks are available. It is possible that the principal could, by imposing less risk, motivate the agent more efficiently if the agent were induced to devote effort to only one task. Since the risk-averse agent must be compensated for bearing risk, the principal may be better off imposing risk related to just one task.

The propositions in the remainder of this section describe situations in which a boundary solution or an interior solution will be optimal, and characterize interior solutions. Before stating the propositions, a simple example will be used to introduce the issues.

Suppose there are two independent and identical tasks, whose outcomes are represented by X_1 and X_2. Suppose further that the agent's action space is $\{(2a^*,0),(0,2a^*),(a^*,a^*),(a^*,0),(0,a^*),(0,0)\}$, where an effort level of 0 represents the minimal effort the agent will exert. Suppose that the probabilities of X_i given a are:

	Probabilities given that		
X_i	$a=2a^*$	$a=a^*$	$a=0$
\$1.10	11/12	1/2	1/12
-.10	1/12	1/2	11/12
$E(X_i \mid a)$	1.00	0.50	0.00
$Var(X_i \mid a)$	0.11	0.36	0.11

The joint outcomes occur with the following probabilities:

		Probabilities given that					
		$a_1=2a^*$	$a_1=0$	$a_1=a^*$	$a_1=a^*$	$a_1=0$	$a_1=0$
Reward	(X_1,X_2)	$a_2=0$	$a_2=2a^*$	$a_2=a^*$	$a_2=0$	$a_2=a^*$	$a_2=0$
s_1	(1.1,1.1)	11/144	11/144	1/4	1/24	1/24	1/144
s_2	(1.1,-.1)	121/144	1/144	1/4	11/24	1/24	11/144
s_3	(-.1,1.1)	1/144	121/144	1/4	1/24	11/24	11/144
s_4	(-.1,-.1)	11/144	11/144	1/4	11/24	11/24	121/144
	$E(X_1+X_2\mid\underline{a})$	1.00	1.00	1.00	0.50	0.50	0.0
	$Var(X_1+X_2\mid\underline{a})$	0.22	0.22	0.72	0.47	0.47	0.22

Let $\underline{s} = (s_1,s_2,s_3,s_4)$. Suppose the principal's problem is:

$$\text{Maximize } E(250X_1 + 250X_2) - E(\underline{s})$$

$$\text{subject to } EU(\underline{s}) - V(a_1 + a_2) \geq \bar{u}$$

$$(a_1,a_2) \text{ maximizes } \{EU(\underline{s}) - V(a_1+a_2)\} .$$

Let $U(s_i) = \sqrt{s_i}$ and $\bar{u} = 10$, and $a^* = 1$. The optimal solution is for the principal to induce the agent to exert $2a^*$ at one task, with the reward for the one task as follows:

$$\bar{s} = 148.84 \text{ and } s_0 = 96.04,$$

where $\bar{s}$ is paid if the outcome is 1.1, and s_0 is paid otherwise. If the principal desired to induce the agent to exert a^* at both of the tasks, the following sharing rule would be optimal:

$$s_1' = 207.40,\ s_2' = s_3' = 144,\ s_4' = 92.16.$$

Looking at the variance as a measure of risk, we note that the outcome is riskier when $\underline{a} = (a_1^*,a_2^*)$ than when $\underline{a} = (2a^*,0)$. However, this risk is not directly of concern to either the principal or the agent, because the principal is risk neutral and the agent is not concerned about the riskiness of the outcomes per se, but rather about the effects on the compensation received. In the example above, $Var(\underline{s}) = 213.0$ while $Var(\underline{s}') = 1667.2$, and $E(\underline{s}) = 144.44$, which is less than $E(\underline{s}') = 146.88$. The principal can thus motivate the agent more efficiently with a boundary solution rather than with an interior solution. In this case, the principal's expected payments to the agent are lower for the sharing rule which imposes less risk (as measured by the variance) on the agent.

Although the variances of the outcomes are not directly of concern to either the principal or the agent, they are indirectly of concern. X_1 and X_2 are not only outcomes, but also signals about the agent's efforts; as such, they provide information about the efforts. The relative magnitudes of the variances of the outcomes are potential surrogates for measures of informativeness, since the variances indicate how the signals (information) about the efforts will vary as the efforts vary. In the example above, a total effort level of 2a* will provide the same total expected outcome, regardless of whether a* is devoted to each of two tasks, or 2a* is devoted to a single task. However, the variance of the outcome is smaller when 2a* is devoted to a single task than when the effort is allocated to two tasks. Since the expected outcomes are the same, the risk-neutral principal desires to allocate effort in the way that provides the most information about shirking. That is, information issues become dominant in the principal's choice of the allocation of effort.

A situation similar to the discrete outcome example above occurs when the X_i's are independent and identically distributed with a normal distribution with mean ka and variance σ^2. If effort a* is devoted to each of two independent tasks, then $E(X_1+X_2|a_1=a^*,a_2=a^*) = 2ka^*$ and $Var(X_1+X_2|a_1=a^*,a_2=a^*) = 2\sigma^2$. If effort 2a* is devoted to just one task, say the first task, then the expected outcome is 2ka*, which is equal to $E(X_1+X_2|a_1=a^*,a_2=a^*)$. However, if the agent is compensated only on the basis of X_1, the corresponding variance is σ^2, which is strictly less than $Var(X_1+X_2|a_1=a^*,a_2=a^*)$. In this situation, then, we might conjecture that a boundary solution is optimal.

The two examples above had $2E(X_i|a^*) = E(X_i|2a^*)$. Clearly, this can hold for all effort levels only when the means are linear in effort. The examples also had

$$Var(X_1+X_2|a_1=a^*,a_2=a^*) > Var(X_1|a_1=2a^*). \tag{3.5.4}$$

Thus, one might conjecture that a boundary solution is optimal in cases where (3.5.4) holds and there are independent and identically distributed outcomes, with the means proportional to effort. It should be pointed out, however, that the additivity of effort would also be critical for this result.

If the X_i's have Poisson distribution with $E(X_i|a_i=a) = ka = Var(X_i|a_i=a)$, then the variances change as the efforts change. If a^* is exerted at each of two independent tasks, then $E(X_1+X_2|a_1=a^*,a_2=a^*) = 2ka^* = Var(X_1+X_2|a_1=a^*,a_2=a^*)$. If $2a^*$ is exerted at one task, say task one, then $E(X_1|a_1=2a^*) = 2ka^* = E(X_1+X_2|a_1=a^*,a_2=a^*) = Var(X_1|a_1=2a^*)$. Therefore, we might expect that the principal would be indifferent between a boundary solution and an interior one.

Finally, consider the exponential distribution, where $E(X_i|a_i=a) = ka$ and $Var(X_i|a_i=a) = k^2a^2$. If a^* is exerted at each of two independent tasks, then $E(X_1+X_2|a_1=a^*,a_2=a^*) = 2ka^*$ and $Var(X_1+X_2|a_1=a^*,a_2=a^*) = 2k^2a^{*2}$. If $2a^*$ is exerted at one task, then $E(X_1|a_1=2a^*) = 2ka^* = E(X_1+X_2|a_1=a^*,a_2=a^*)$ but $Var(X_1|a_1=2a^*) = 4k^2a^{*2} > Var(X_1+X_2|a_1=a^*,a_2=a^*)$. Thus, in this situation, we might conjecture that an interior solution, rather than a boundary solution, would be optimal.

The propositions below substantiate the intuitive arguments above concerning when an interior solution or a boundary solution is optimal, given that the expected outcomes of independent and identical tasks are proportional to effort expended. If the expected outcomes are nonlinear in effort, then the situations become more complicated.

Initially, the normal distribution with constant variance but with mean a function of effort will be considered. This case is of particular interest, since it is the only distribution in Q (see (2.7)) whose variance is independent of the agent's effort. The following proposition states conditions under which a boundary solution is optimal.

Proposition 3.5.1. Suppose the principal is risk neutral, $U(s) = 2\sqrt{s}$, and x_1 and x_2 are conditionally independent and identically distributed normally with mean ka and constant variance. Suppose further that $V(\underline{a}) = V(\Sigma a_i)$. Then a boundary solution is optimal.[4]

The proposition below characterizes optimal unique interior solutions.

Proposition 3.5.2. Suppose the principal is risk neutral, the agent is risk averse, and $g(\underline{x}|\underline{a}) = f(x_1|a_1)f(x_2|a_2)$, i.e., x_1 and x_2 are conditionally independent and identically distributed. Suppose further that $V(\underline{a}) = V(\Sigma a_i)$. If a unique interior solution is optimal, then the optimal solution has $a_1^* = a_2^*$ and $\mu_1^* = \mu_2^*$, where μ_1 and μ_2 are the Lagrangian multipliers described earlier.

This result is independent of the utility function of the risk-averse agent or the distribution of x_i given a_i; the critical element is that the outcomes are conditionally independent and identically distributed. This result does not say that all agency problems such that the principal is risk neutral, the agent is risk averse, and the outcomes are conditionally independent and identically distributed have solutions of $a_1^* = a_2^*$ and $\mu_1^* = \mu_2^*$; this is evident from Proposition 3.5.1. Proposition 3.5.2 indicates that if the optimal solution has the agent allocating nonzero effort to each task, then the efforts should be equal at each task if the tasks present independent and identical expected returns to the principal. The following proposition, which applies to the one-parameter exponential family (see (2.7)), describes conditions under which an interior solution is optimal. These conditions are sufficient but not necessary.

Proposition 3.5.3. Suppose the principal is risk neutral, $U(s) = 2\sqrt{s}$, and $g(\underline{x}|\underline{a}) = f(x_1|a_1)f(x_2|a_2)$, where $f(\cdot|a)$ belongs to Q and has mean $M(a)$, where $M(0) \geq 0$ and $M'(a) > 0$. Suppose further that $V(\underline{a}) = V(\Sigma a_i)$. Let a^* be the optimal effort in the one-task problem.

(i) If $M(a)$ is concave and

$$z'(a^*)M'(a^*)/[z'(a^*/2)M'(a^*/2)] < 1/2, \qquad (3.5.5)$$

then a boundary solution is not optimal.

(ii) If $M(a)$ is strictly concave and

$$z'(a^*)M'(a^*)/[z'(a^*/2)M'(a^*/2)] \leq 1/2, \qquad (3.5.6)$$

then a boundary solution is not optimal. In both cases, if a unique interior solution is optimal, then $a_1^* = a_2^*$ and $\mu_1^* = \mu_2^*$.

As shown in below in Corollary 3.5.4, $z'(a)/z'(a/2)$ is often independent of a, and hence one need not actually solve for the optimal one-task effort.

Corollary 3.5.4. Under the conditions in Proposition 3.5.3 if $M(a) = ka$ and $z'(a^*)/z'(a^*/2) < 1/2$, then an interior solution is optimal. In particular,

(i) For the exponential distribution with parameter $1/(ka)$, an interior solution is optimal $(z'(a)/z'(a/2) = 1/4)$.

(ii) For the gamma distribution with parameters $n/(ka)$ and n, an interior solution is optimal $(z'(a)/z'(a/2) = 1/4)$.

The following cases do not satisfy (3.5.5) but are included for purposes of comparison:

(iii) The Poisson distribution with mean ka has $z'(a)/z'(a/2) = 1/2$.

(iv) The normal distribution with mean ka and constant variance has $z'(a)/z'(a/2) = 1$.

The normal distribution should not, of course, satisfy (3.5.5) in view of Proposition 3.5.1.

In each of the cases in Corollary 3.5.4 the expected outcomes increase linearly with the agent's efforts. In case (iii), the variances of the outcomes also increase linearly with the agent's efforts. In case (iv), the variances of the outcomes are unaffected by the efforts. In cases (i) and (ii), the variances of the outcomes increase quadratically with the efforts. A boundary solution is optimal in case (iv), where the rate of increase in the variance is strictly less than the rate of increase in the mean. An interior solution is optimal in cases (i) and (ii), where the rates of increase in the variances are strictly greater than the rates of increases in the means.

The following two propositions characterize optimal interior solutions when the means of the outcomes are linear in effort.

Proposition 3.5.5. Suppose the principal is risk neutral, $U(s) = 2\sqrt{s}$, and $g(\underline{x}|\underline{a}) = f(x_1|a_1)f(x_2|a_2)$, where $f(\cdot|a)$ belongs to Q and has mean M(a). Suppose further that $V(\underline{a}) = V(\Sigma a_i)$. If $M(a) = ka$ and $z''(a)/z'^2(a)$ is strictly monotonic, then an optimal interior solution is unique and has $a_1^* = a_2^*$ and $\mu_1^* = \mu_2^*$. The strict monotonicity is satisfied by the exponential and gamma distributions (given that $M(a) = ka$), but not by the normal or Poisson distributions.

Proposition 3.5.6. Suppose the principal is risk neutral, $U(s) = 2\sqrt{s}$, and $g(\underline{x}|\underline{a}) = f(x_1|a_1)f(x_2|a_2)$, where $f(\cdot|a)$ has mean M(a). Suppose further that $V(\underline{a}) = V(\Sigma a_i)$. If $M(a) = ka$ and $I'(a)/I^2(a)$ is strictly monotonic, where $I(a) = \int f_a^2/f \, dx$, then an optimal interior solution is unique and has $a_1^* = a_2^*$ and $\mu_1^* = \mu_2^*$.

I(a) is called Fisher's information about a contained in x, and is a useful concept in mathematical statistics (see, e.g., Cox and Hinkley, 1974).

The next corollary demonstrates that in part, the shape of the expected outcome function determines whether the optimal solution will be interior.

Corollary 3.5.7. Under the conditions in Proposition 3.5.3 if $M(a) = a^{\alpha}$, then an interior solution is optimal if $f(\cdot|a)$ is

(i) Normal $(M(a),\sigma^2)$ and $0 < \alpha \leq 1/2$ or

(ii) Exponential $(1/M(a))$ and $0 < \alpha < 1$ or

(iii) Poisson $(M(a))$ and $0 < \alpha < 1$.

It is well known that knowing f_a/f is equivalent to knowing the likelihood of a given the observations. For the exponential family Q, f_a/f is given by $z'(a)(x-M(a))$, where M(a) is the mean of x conditional on a. It is z'(a) and M(a) which play an important role in determining whether a boundary solution or an interior solution is optimal. This might be expected, for the x_i's are not only outcomes, but also signals about the efforts that have been expended. Since f_a/f is sufficient for the likelihood of a given x, z'(a) and M(a) together measure, to a certain degree, the informativeness of x about a.

It is interesting to compare the results for the second best case with those for the first best case. In the first best case, (3.5.1) indicates that if M(a) = ka, then whatever the distribution of x given a, the principal will be indifferent between an interior solution or a boundary one, as long as the total amount of effort expended is the same in both cases. In the second best case, however, Proposition 3.5.1 says that if the distribution is normal with mean ka and constant variance, then a boundary solution is optimal. On the other hand, if the distribution is exponential or gamma with mean ka, then an interior solution is optimal (Corollary 3.5.4).

If, in the first best case, the means are concave in effort, then (3.5.1) indicates that an interior solution is optimal, and the optimal efforts are equal if the mean functions are identical ($M'(a_1^*) = M'(a_2^*)$ implies that $a_1^* = a_2^*$). Corollary 3.5.7 indicates that for a specific second best case with concave means, a similar result concerning the optimality of an interior solution holds.

The results up to this point have assumed that x_1 and x_2 are conditionally independent and identically distributed. The next two propositions deal with the case of conditionally independent but nonidentically distributed x_i's.

Proposition 3.5.8. Suppose the principal is risk neutral, $U(s) = 2\sqrt{s}$, and $g(\underline{x}|\underline{a}) = f(x_1|a_1)h(x_2|a_2)$, where $f(.|a_1)$ and $h(.|a_2)$ belong to Q and $E(x_i|a=a_i) = k_i a_i$. Suppose further that $V(\underline{a}) = V(\Sigma a_i)$.

(i) If x_i has an exponential distribution with mean $k_i a_i$, then $k_1 > k_2$ implies that $a_1^* > a_2^*$ and $\mu_1^* > \mu_2^*$.

(ii) If x_i has a gamma distribution with mean $k_i a_i$, then $k_1 > k_2$ implies that $a_1^* > a_2^*$ and $\mu_1^* > \mu_2^*$.

(iii) If x_i has a normal distribution with mean $k_i a_i$ and constant variance, then $k_1 > k_2$ implies that the optimal solution is a boundary solution, with $a_1^* > 0$ and $a_2^* = 0$.

(iv) If x_i has a Poisson distribution with mean $k_i a_i$, then $k_1 > k_2$ implies that the optimal solution is a boundary solution with $a_1^* > 0$ and $a_2^* = 0$.

The following proposition states that at least for a specific second best case, the optimal Lagrangian multipliers are positive. Signing the multipliers is of importance because of their critical role in the determination of the optimal sharing rule. For example, if the density of x_i given a_i satisfies the monotone likelihood ratio property in x_i for all i, then the positivity of the μ_i's guarantees that the optimal sharing rule is increasing in each x_i. It can be shown that under certain conditions in the second best case, not all the μ_i's can be zero, and hence in the situations above where the optimal multipliers μ_i^* are equal, they must be positive.

Proposition 3.5.9. Suppose the principal is risk neutral, $U(s) = 2\sqrt{s}$, and $g(\underline{x}|\underline{a}) = f(x_1|a_1)h(x_2|a_2)$, where $f(\cdot|a_1)$ and $h(\cdot|a_2)$ belong to Q. Suppose further that $V(\underline{a}) = V(\Sigma a_i)$. If an interior solution is optimal, then $\mu_1^* > 0$ and $\mu_2^* > 0$.

Note that in Proposition 3.5.9, x_1 and x_2 need not be identically distributed, although they are conditionally independent. Furthermore, the result holds for general $V(\underline{a})$, as long as $\partial V/\partial a_i > 0$, i=1,2.

3.6 APPLICATION TO SALES FORCE MANAGEMENT

In this section, the previous analysis of multidimensional effort situations is applied to the problem of sales force management. Steinbrink (1978) depicts the critical role of compensation of a sales force as follows:

> Any discussion with sales executives would bring forth a consensus that compensation is the most important element in a program for the management and motivation of a field sales force. It can also be the most complex.
>
> Consider the job of salespeople in the field. They face direct and aggressive competition daily. Rejection by customers and prospects is a constant negative force. Success in selling demands a high degree of self-discipline, persistence, and enthusiasm. As a result, salespeople need extraordinary encouragement, incentive and motivation in order to function effectively.
>
> . . .A properly designed and implemented compensation plan must be geared to the needs of the company and to the products or services the company sells. At the same time, it must attract good salesmen in the first place . . .

Management of the sales force has been the focus of a great deal of research, much of it empirical. Steinbrink (1978), in a survey of 380 companies across 34 industries, found that most companies favored a combination of salary, commission, and bonus schemes. Typical commissions used were

1) Fixed commissions on all sales

2) Different rates by product category

3) On sales above a determined goal

4) On product gross margin.

These commission schemes are all examples of linear sharing rules.

Farley (1964), Berger (1975), and Weinberg (1975, 1978) studied the problem of "jointly optimal" compensation schemes. They assumed a given compensation system (a commission scheme based on gross margin) and sought to determine if that system is incentive compatible, meaning that the salesperson will be induced to choose levels of effort which the company desires. In these analyses, the measure of effort is taken to be time spent selling. The total time available is assumed to be fixed and the decisions are how to allocate the total time across several products.

Farley (1964) demonstrated that if a commission system based on gross margin is used, the commission rates should be the same for all products in the case where both the firm and the salespeople are income maximizers. Weinberg (1975) extended Farley's result to include the choice of discounts on each product as well as the choice of time spent selling each product. Both papers assume that the time spent selling one product does not affect the sales of any other product. Furthermore, sales are considered to be a deterministic function of time, although the conclusions are unaffected by uncertainty because of the assumed risk neutrality of both the firm and the salespeople. Weinberg (1978) maintained the assumption of risk neutrality of the salespeople, and further extended his and Farley's analyses by allowing for interdependence of product sales and relaxing the assumption that salespeople maximize income. Even in these situations, an equal gross margin commission system is incentive compatible if the firm's objective is to maximize expected gross profits.

Berger (1975) examined the combined effects of uncertainty and non-neutral risk attitudes on the part of the salespeople. He retained Weinberg's and Farley's assumption of constant marginal cost per product, but treated sales of each product as a random variable parameterized by the time spent selling that product. Berger demonstrated that if a commission scheme is used in this situation, it may be undesirable for the firm to set equal commission rates for all products.

The agency model allows for many of the important factors in the sales force managment problem and provides a more complete analysis of the problem by determining an incentive scheme which motivates the salesperson to make decisions that are Pareto optimal, rather than taking the compensation scheme as given. The problem will therefore be examined below in an agency framework.[5] Interdependence of products, provision of enough net benefits for the salesperson to join and stay with the firm, and also the salesperson's tradeoff between money and time spent selling are incorporated. In connection with this, the total time spent selling in a given time period will be a choice variable. In order to focus on the motivational rather than risk sharing aspects of the problem, it will be assumed that the salesperson is risk averse, and the firm is risk neutral and therefore desires to maximize expected profit. The agency theory analysis isolates conditions under which some sort of commission scheme is Pareto optimal, and shows that even when a commission scheme is Pareto optimal, the commissions are generally unequal.

In this analysis, effort will be interpreted as "time spent selling," and n will represent the number of products available to be sold. It will be assumed that the salesperson has no intrinsic disutility for selling any particular product, so that the disutility function may be taken to be $V(\Sigma a_i)$, with V increasing and convex. The remaining notation will be as defined previously, with x_i denoting the difference between sales revenue and variable noncompensation costs for product i.[6]

Suppose that the principal (firm) and the agent (salesperson) have identical beliefs. This might be the case when the salesperson is first hired. Suppose also that the firm and the salesperson are in a first best situation, either because they are acting cooperatively or because the firm can perfectly observe the times spent selling each product. Recall that in the first best situation, it makes no difference whether the principal observes only the total outcome, x, or the vector of outcomes, $\underline{x}$. Since the firm is risk neutral and the salesperson is risk averse, the optimal sharing rules is a constant salary $c = U'^{-1}(1/\lambda)$ for the salesperson, with the firm receiving the remainder, x-c. The firm requires that the salesperson choose sales effort so that

$$\partial E(x|a)/\partial a_i = \partial E(x|a)/\partial a_j, \quad i,j=1,\ldots,n. \tag{3.6.1}$$

The interpretation in the cooperative setting is that the salesperson happily supplies effort levels a_i satisfying (3.6.1) in return for the salary c, since in doing so, he or she receives the market utility, $\bar{u}$. In the perfect observability setting, the firm pays the salesperson a salary c if effort levels a_i satisfying (3.6.1) are exerted, and pays nothing otherwise. Observe that in the first best case, the salesperson chooses effort levels according to their effect on mean outcome.

Suppose $E(x_i) = M_i c_i a_i$, where M_i represents the contribution margin (sales revenues minus variable noncompensation costs) per unit of product i, and $c_i a_i$ represents the expected quantity of product i that will be sold if effort a_i is exerted. The analysis in Section 3.5 then applies. If $M_1 c_1 = M_2 c_2$, i.e., if the contributions per unit of time spent selling are equal, the total effort expended is the only concern. If $M_1 c_1 > M_2 c_2$, then the Pareto optimal strategy is for the agent to devote effort only to the first product. Under a more general return structure, the efforts a_1 and a_2 will be nonzero and unequal. If the mean functions are identical and nonlinear and monotone in a_i, then the optimal efforts are such that $a_1 = a_2$.

In practice, a straight salary is seldom used for salespeople because of imperfect observability and imperfect cooperation (moral hazard). In such situations, a second best analysis is appropriate. Consider first the one-product case, in which the interior portion of the sharing rule is characterized by

$$\frac{1}{U'(s(x))} = \lambda + \mu \frac{g_a(x|a)}{g(x|a)} ,$$

where x and a are univariate and the subscript a denotes differentiation with respect to a. Examples of specific sharing rules are provided in Table I for two members of the HARA class of utility functions and two members of the one-parameter exponential family Q.

Sharing Rule Given $g(x|a) =$

$U(s)$	$(M(a))^{-1}\exp[-x/M(a)]$, $M'(a) > 0$	$(2\pi\sigma)^{-1}\exp[-(x-M(a))^2/2\sigma^2]$, $M'(a)>0$
$\ln s$	$s_e = \lambda + \frac{\mu M'(a^*)}{M^2(a^*)}(x-M(a^*))$	$s_N = \lambda - \frac{\mu M(a^*)M'(a^*)}{\sigma^2} + \frac{\mu M'(a^*)}{\sigma^2}x$
$s^{1/b}, b>1$	$(\frac{s_e}{b})^{b/(b-1)}$	$(\frac{s_N}{b})^{b/(b-1)}$

Table I. Examples in One-Product Case

Observe that when $U(s) = \ln s$, the sharing rules shown (and others corresponding to different members of Q, the one-parameter exponential family) can be interpreted as a salary plus commission on the outcome x, a scheme commonly found in practice. If the agent's utility function is a concave power function, then the resulting sharing rule is a convex power function of a linear form. The compensation schemes which pay a salary plus bonus commissions (e.g., $s(x) = m + m_1 x$ if $x < x_0$, $s(x) = m + m_1 x + m_2(x-x_0)$ if $x > x_0$) can be considered as approximations to these sharing rules.

The case where $n > 1$ is more complicated if the agent's utility function is a power function, since cross terms in the x_i's appear. In order to examine conditions under which it is optimal to use a salary-plus-commission scheme, the agent's utility function will be taken to be $U(s) = \ln s$, since this is the only situation in which a linear scheme can be optimal (see Section 3.4). The examples below employ the normal distribution because of its convenient representation for dependent random variables. For purposes of illustration, it suffices to take $n = 2$.

Suppose then that $U(s) = \ln s$, $n = 2$, and that $\underline{x} \sim N(\theta(\underline{a}), \Sigma(\underline{a}))$, where $\theta(\underline{a}) = (\theta_1(\underline{a}), \theta_2(\underline{a}))$ and $\Sigma(\underline{a})$ is the covariance matrix

$$\begin{pmatrix} \sigma_1^2(\underline{a}) & \rho(\underline{a})\,\sigma_1(\underline{a})\,\sigma_2(\underline{a}) \\ \rho(\underline{a})\,\sigma_1(\underline{a})\,\sigma_2(\underline{a}) & \sigma_2^2(\underline{a}) \end{pmatrix}.$$

At this level of generality, the optimal sharing rule is quite complicated (see Appendix 2). It is not separable in x_1 and x_2, and therefore is not a salary plus commission scheme. A correlation coefficient which is constant (independent of $\underline{a}$) is <u>not</u> sufficient for the sharing rule to be a salary plus commission scheme, although $\rho = 0$ does lead to a sharing rule which is additively separable in x_1 and x_2. Sufficient conditions for a salary plus commission scheme to be optimal are that both the correlation coefficient and the variances be constant, with $\rho^2 \neq 1$. The commissions are determined by ρ, marginal increases in the means $\theta_i(a)$ at a^*, the variances σ_i^2 and the multipliers μ_i.

Three especially interesting results of the example above are:

(1) In the case of the normal distribution with log utility, independence of the products is enough to guarantee additive separability (in x_1 and x_2) of the optimal sharing rule, but is not enough to guarantee that the optimal

sharing rule will be a linear sharing rule. That is, the optimal sharing rule is not a (salary plus) commission scheme, let alone an equal commission rate scheme.

(2) It is not necessary for the products to be independent in order for a salary plus commission scheme to be optimal, or for a separable sharing rule to be optimal.

(3) The optimal commission rates are generally not equal across products.

The agency analysis applied to the sales force management problem indicates that only under very special circumstances is a commission scheme Pareto optimal. In practice, of course, commission schemes are favored because of their simplicity and ease of application, as well as their recognized incentive effects. If commission rates are used with risk averse salespeople who face uncertainty in sales, the rates should most likely not be equal across products, according to the analysis above.

The results in Section 3.5 on the allocation of additive effort with independent outcomes provide some further insights about optimal compensation schemes for salespeople. It should be recalled that most of the results in Section 3.5 were proved only for $U(s) = 2\sqrt{s}$. Thus, the remarks that follow are restricted by the assumption of that particular utility function for wealth for the agent.

A principle commonly taught in managerial accounting texts is that under certainty, in order to maximize profits given one scarce factor of production, a firm should manufacture the product which returns the highest contribution margin per unit of the scarce factor (see, e.g., Horngren (1982, p. 373)). This principle does not necessarily hold in the agency setting. In the first best case, if the means are linear in effort, then the principle holds. In addition, Proposition 3.5.8 indicates that in the second best case, if expected returns are linear in effort, then all the agent's effort should be put into selling the product with the highest expected return per unit of effort if the underlying distribution is normal with constant variance, or Poisson. However, if the underlying distribution is exponential, then more effort should be put into selling the product with the higher expected return per unit of effort, but both efforts will be positive unless the expected returns per unit of effort are very different. (See the discussion after Proposition 3.5.8.)

For the exponential distribution with $E(x_i|a_i) = k_i a_i$, the optimal sharing rule is given by

$$s(x_1,x_2) = [\lambda + \frac{\mu_1^*}{k_1 a_1^{*2}}(x_1 - k_1 a_1^*) + \frac{\mu_2}{k_2 a_2^{*2}}(x_2 - k_2 a_2^*)]^2.$$

If $k_1 > k_2$, then $\mu_1^* > \mu_2^*$ and $a_1^* > a_2^*$. Equation [3] in the proof of Proposition 3.5.8 shows that $\mu_1^*/a_1^{*2} = \mu_2^*/a_2^{*2}$. Therefore, $\mu_1^*/(k_1 a_1^{*2}) < \mu_2^*/(k_2 a_2^{*2})$. This implies that when k_1 is <u>greater</u> than k_2 (the expected return per unit of effort is <u>greater</u> for product one than for product two), the agent's compensation per unit of x_1 (the return on product one) is <u>less</u> than the compensation per unit of x_2.

Continuing with the exponential distribution case, if $k_1 = k_2$, then $a_1^* = a_2^*$ and $\mu_1^* = \mu_2^*$ (Proposition 3.5.5). The agent's compensation per unit of x_1 is equal to the compensation per unit of x_2, and the sharing rule can be written as

$$s(x_1,x_2) = [\ \lambda + \frac{\mu_1^*}{k_1 a_1^{*2}}(x_1 + x_2) - \frac{2\mu_1^*}{a_1^*}\ ]^2.$$

Thus, the information (x_1,x_2) has no value in addition to $x_1 + x_2$. A similar result holds for more general situations, also. Proposition 3.5.2 says that if the principal is risk neutral, the agent is risk averse, $V(\underline{a}) = V(\Sigma a_i)$, the x_i's given a_i are independent and identically distributed, and a unique interior solution ($a_1^* > 0$, $a_2^* > 0$) is optimal, then $a_1^* = a_2^*$ and $\mu_1^* =$

$\frac{1}{2}$. Under these conditions, the agent's compensation per unit of x_1 is equal to the compensation per unit of x_2. It is important to note that if x_i given a_i has a normal distribution with mean ka_i and variance σ^2, and the x_i's given a_i are independent, a boundary solution (e.g., $a_1^* > 0$, $a_2^* = 0$) is optimal. In this case, the agent would receive no compensation based on x_2.

Up to this point, the focus has been on a single agent exerting multiple efforts. A related topic is that of multiple agents, which is pertinent here because a firm will generally have more than one salesperson. Feltham (1977b) examined the use of penalty contracts when all the agents are identical, and Holmstrom (1982) showed that the effectiveness of group penalties will be hampered by limited endowments of the agents, especially as the number of agents becomes large.

An important question in the multiple agent problem is whether or not each agent should be rewarded independently of the others' performances. Holmstrom (1982) showed that if the agents' outcomes are correlated with each other through the common uncertainty they face, then basing agent i's share on each agent's outcome helps reduce the uncontrollable randomness in agent i's reward. Holmstrom (1982, p. 335) stated that

> . . .forcing agents to compete with each other is valueless if there is no common underlying uncertainty. In this setting, the benefits from competition itself are nil. What is of value is the information that may be gained from peer performance. Competition among agents is a consequence of attempts to exploit this information.

Only aggregate information about peer performance is used in the optimal sharing rules if the aggregate measure captures all the relevant information about the common uncertainty. Of course, if the agents' outcomes are independent of one another, then the optimal sharing rule for agent i depends only on agent i's outcome.

One of the traditional principles in performance evaluation within the firm is the principle that a person should be held responsible only for those factors (e.g., costs or revenues) over which he or she has control. Basing the sharing rule for agent i only on agent i's outcome is clearly consistent with the controllability principle. Basing the sharing rule for agent i on the outcomes of other agents when there is common uncertainty is, at first glance, inconsistent with the controllability principle. However, the reason that the compensation for each agent may depend on the outcomes of other agents is that the principal can gain information about the random state, and hence gain information about the efforts expended by each agent. That is, the principal can gain information about each agent's input (effort), over which the agent has direct control. Thus, there is no conflict with the controllability principle in this case. The apparent conflict occurs because the focus of the controllability principle has been transferred from outputs to inputs (cf. Baiman (1982, pp. 197-198)).

The last modification to the standard agency analysis for the problem of sales force management relates to noneffort decisions. Frequently, the salesperson must not only make several effort decisions, but also make several "risk" decisions which do not require expenditures of effort. The choices of discounts to offer on each product are examples of such risk decisions. Weinberg (1975, p. 938) identifies the following situations in which an agent might have control over the price:

> (1) perishable agricultural products; . . . (2) sales involving trade-ins in which the salesman has control over the evaluation of the trade-in, e.g., automobiles; (3) systems selling in which the salesman has a wide range of latitude in specifying the combination of services to be provided, e.g., contractors and consultants; (4) some retail situations in which the local store manager has control over price of at least some of the items sold in his store; (5) liquidation sales of obsolete product lines or retailer distress sales; and (6) highly competitive markets in which customers are price bargainers . . . .

One approach to the problem of incorporating both risk and effort decisions was taken by Itami (1979), who examined optimal linear goal-based incentive schemes under uncertainty. In his model, a risk decision is made by the agent before the state of nature is observed. The agent then chooses an effort level based on the risk decision and the observed state of nature, resulting in a deterministic output which is a function of the agent's two decisions and the state of nature. For example, the divisional manager of a large corporation might make investment decisions on projects before the environmental conditions are revealed. The effort expended and the known state then determine the output.

The simplest agency theory approach to the problem of incorporating both risk and effort decisions is to assume that both of the agent's decisions are made before the state of nature (or any other information) is observed. This approach will now be briefly discussed. The form of the optimal sharing rule is derived rather than assumed. Furthermore, because risk-sharing aspects are important in this setting, both the principal and the agent are assumed to be risk averse.

As Itami points out, there is a direct and an indirect effect of the agent's effort on his or her utility, while there is only an indirect effect from the risk decisions. Up to this point, it has been assumed that the agent's utility is separable in effort and wealth. This assumption leads to a characterization of the optimal sharing rule that is independent of the agent's disutility for effort, although the indirect effects of effort expended are captured via the terms g_{a_j}/g. The more general utility function $U(s(\underline{x}),\underline{a})$ for the agent leads to a characterization of the optimal sharing rule that captures both the direct and indirect effects of the agent's effort. When there are no risk decisions, optimality requires that for interior solutions,

$$\frac{W'(x-s(\underline{x}))}{U_s(s(\underline{x}),\underline{a})} = \lambda + \sum_j \mu_j \left[\frac{U_{a_j s}(s(\underline{x}),\underline{a})}{U_s(s(\underline{x}),\underline{a})} + \frac{g_{a_j}(\underline{x}|\underline{a})}{g(\underline{x}|\underline{a})} \right] ,$$

where the subscripts a_j on U and g denote differentiation with respect to a_j, and the subscripts s denote differentiation with respect to s. The major implication of nonseparability of the utility function is that the role of effort is explicit, as is interaction between effort and compensation. It is still true that if the agent is risk neutral, then the first best solution is achievable by a sharing rule of the form x-k.

Because the important distinguishing feature of effort decisions is their twofold effect on the utility function, the general form of the utility function is used here. Letting r_i denote the risk decision for task i and $\underline{r} = (r_1,\ldots,r_m)$, the principal's problem is[7]

$$\underset{s(\underline{x}),\underline{a},\underline{r}}{\text{Maximize}} \quad \int W(x-s(\underline{x}))g(\underline{x}|\underline{a},\underline{r})\ d\underline{x}$$

$$\text{subject to} \qquad \int U(s(\underline{x}),\underline{a})g(\underline{x}|\underline{a},\underline{r})\ d\underline{x} \geq \bar{u} \qquad :\lambda$$

$$\frac{\partial}{\partial a_j}\int U(s(\underline{x}),\underline{a})g(\underline{x}|\underline{a},\underline{r})d\underline{x} = 0,\ j=1,\ldots,n \qquad :\mu_j$$

$$\frac{\partial}{\partial r_j}\int U(s(\underline{x}),\underline{a})g(\underline{x}|\underline{a},\underline{r})d\underline{x} = 0,\ j=1,\ldots,m. \qquad :\beta_j$$

To the right of the constraints above are their associated multipliers. The interior portion of the optimal sharing rule is characterized by

$$\frac{W'(x-s(\underline{x}))}{U_s(s(\underline{x}),\underline{a})} = \lambda + \sum_j \mu_j\left[\frac{U_{a_js}(\ )}{U_s(\ )} + \frac{g_{a_j}(\ )}{g(\ )}\right] + \sum_j \beta_j \frac{g_{r_j}(\ )}{g(\ )} .$$

It should be noted that there is an implicit assumption that the r_i's do not satisfy first-order stochastic dominance, since otherwise the principal and the agent would agree on the choices of the r_i's and there would be no incentive problem with respect to the r_i's.

Suppose next, as Weinberg (1975) did, that the gross margin generated by sales of product i is given by

$x_i = P_i(1-r_i)Q_i - K_iQ_i$, where

P_i = nominal selling price per unit of product i,

r_i = discount (decimal) on product i,

Q_i = quantity (units) of product i sold,

K_i = variable nonselling cost per unit of product i, and

$M_i = P_i(1-r_i) - K_i$ = gross margin on product i.

Weinberg (1975) sought to determine if an equal-commission scheme is incentive compatible when there both risk and effort decisions. An agency theory analysis suggests that such a scheme is not Pareto optimal. Suppose $Q_i \sim N(\theta_i(a_i,r_i),\sigma_i(r_i))$. Then $x_i \sim N(P_i(1-r_i)-K_i)\theta_i(a_i,r_i)$, $(P_i(1-r_i)-K_i)^2\sigma_i^2(r_i))$. Previous analysis indicates that if both the principal and the agent are risk averse with utility functions in the HARA class, then the optimal sharing rule is in general not additively separable in x_1 and x_2. If the principal is risk neutral and the agent's utility is ln s - V(a), then previous remarks concerning the optimality of a commission scheme in the normal distribution example with no risk decisions apply.

Demski and Sappington (1983) examined the situation in which it is desired to motivate an individual to obtain and use information which is personally costly (in a pecuniary or nonpecuniary sense) for the individual to obtain. Their analysis may provide insights for the sales force management problem when the salesperson has the option or the ability to observe private information before making risk decisions.

3.7 SUMMARY AND DISCUSSION

This chapter derived optimal incentive schemes when the agent has several tasks over which to exert effort, and the principal and the agent have homogeneous beliefs about the outcome distribution. In the first best case, where there is no moral hazard problem, the major issue is risk sharing, and the results are similar in nature to the one-dimensional effort case. If one individual is risk neutral and the other is risk averse, then the risk neutral individual bears all the risk, receiving the uncertain outcome less a constant fee. If both individuals are risk averse, then the risk sharing aspect is prominent; even if the disaggregated information, $x = (x_1,\ldots,x_n)$, is observed, the sharing rule depends only on the sum of the x_i's.

In the analysis of the second best case, where there is a moral hazard problem, the principal was assumed to be risk neutral in order to focus on motivational issues. As in the single-task case, the first best solution is achievable when the agent is risk neutral. When the agent is risk averse, the optimal sharing rule can be as simple as a salary plus commission, or can be more complicated, depending on the distribution of the outcomes and the agent's utility function. In general, it is much more difficult to determine when the sharing rule will be increasing in each outcome, x_i, than in the single-dimensional effort and output case. There are two reasons: the sign of each of the Lagrangian multipliers μ_i must be determined, and the question of multivariate stochastic dominance must be addressed. Each of these problems can be analyzed only in special cases.

The analysis of the value of additional information is also more complicated than in the single-dimensional effort and output case. The applicability of the results of Amershi (1982) for the multidimensional effort case was discussed. The use of additional disaggregated information was demonstrated by means of examples. It was shown that in the case where a salary plus commission scheme is optimal, the commissions related to each task will generally be unequal.

The next question addressed was whether a manager should receive separate rewards for the outcomes from the different tasks. It was shown that a strong form of independence (see (3.4.1)) is neither necessary nor sufficient for an optimal sharing rule to be additively separable in the outcomes. If the principal is risk neutral and the agent's utility function is in the HARA class, then a necessary condition for additive separability of the optimal sharing is that the agent have a log utility function. If the principal and the agent are identically risk averse, with identical log utility functions, then the optimal sharing rule is not additively separable.

The remainder of this chapter focused on situations in which effort is additive, as when effort represents time devoted to different tasks. The optimal sharing rules in the first best and second best situations were examined under various assumptions about the means and distributions of the outcomes. In the additive effort case where there is no intrinsic disutility for any particular task, it is of interest to determine whether the principal can most efficiently induce a risk averse agent to allocate all effort to one task, or to diversify by allocating effort to each task. This section showed that the nature of the outcome distribution is an important factor in determining whether the optimal solution will be boundary (all effort devoted to one task) or interior (effort spread across tasks). The critical factor, however, appears to be the relationship between effort expended and the mean of the distribution. Conditions under which an optimal interior solution is unique were found, and it was shown that if an optimal interior solution is unique, then the optimal efforts for both tasks are equal, as are the Lagrangian multipliers μ_i.

The additive effort results were applied to the sales force management problem. As remarked earier, simple commission schemes are rarely Pareto optimal; even when they are optimal, the commissions are generally not equal across products. However, if the principal is risk neutral, the agent is risk averse, $V(\underline{a}) = V(\Sigma a_i)$, the x_i's given a_i are independent and identically distributed, and a unique interior solution ($a_1^* > 0$, $a_2^* > 0$) is optimal, then the agent's compensation per unit of x_1 is equal to the compensation per unit of x_2. The multiple salesperson firm was briefly discussed, as was the addition of risk decisions (not involving effort) by the salesperson.

It has long been recognized that dysfunctional behavior on the part of managers can be induced by their focus on short-term personal goals rather than long-term company goals. Moreover, the company may unwittingly pressure managers to make decisions which will increase short-term profits at the expense of long-term goals. One of the obvious aspects of a solution is to extend the performance evaluation period from, for example, one year to several years. A brief comparison of the allocation of effort problem and a multiperiod problem follows.

Consider the situation in which the agent chooses one action a_i in each of n time periods, resulting in monetary outcomes x_i which are observed by both the principal and the agent at the end of period i. The agent's action in any period and the sharing rule for each period can then depend on the outcomes from previous periods. For ease of exposition, the two-period horizon will be considered here. The principal's utility for the two-period horizon is now $W(x_1-s_1(x_1),x_2-s_2(x_1,x_2))$, and the agent's utility is $U(s_1(x_1),s_2(x_1,x_2),a_1,a_2(x_1))$. Let $g(x_1,x_2,a_1,a_2(x_1)) = h(x_2|a_1,x_1,a_2(x_1))f(x_1|a_1)$. The principal's problem is then

$$\underset{a_i,s_i}{\text{Maximize}} \iint W(x_1-s_1(x_1),x_2-s_2(x_1,x_2))g(x_1,x_2,a_1,a_2(x_1))dx_1dx_2 \qquad (3.7.1)$$

subject to

$$\iint U(s_1(x_1),a_1,s_2(x_1,x_2),a_2(x_1))g(x_1,x_2,a_1,a_2(x_1))dx_1dx_2 \geq \bar{u} \qquad (3.7.2)$$

and a_1 and $a_2(\cdot)$ maximize the left-hand side of (3.7.2).

Let E_2 denote expectation with respect to $h(\cdot)$. The optimal sharing rules are characterized by

$$\frac{E_2 W_{s_1}}{E_2 U_{s_1}} = \lambda + \mu_1 \frac{f_{a_1}(x_1|a_1)}{f(x_1|a_1)} + \mu_1 \frac{(E_2 U_{s_1})_{a_1}}{E_2 U_{s_1}}$$

$$+ \mu_2(x_1) \frac{(E_2 U_{s_1})_{a_2}}{E_2 U_{s_1}}, \text{ for almost every } x_1,$$

and

$$\frac{W_{s_2}}{U_{s_2}} = \lambda + \mu_1 \frac{g_{a_1}(\cdot)}{g(\cdot)} + \mu_2(x_1) \frac{h_{a_2}(\cdot)}{h(\cdot)}, \text{ for almost every } (x_1, x_2).$$

Two special cases are of interest. The first is the case in which the principal's and agent's utilities are additive over time, with discount factors β and α, respectively. Suppose the principal and the agent agree on the contract at the beginning of the two-period horizon and each individual is committed to the contract for the entire time horizon. Then the principal's expected utility is

$$\int W(x_1-s_1(x_1))f(x_1|a_1)dx_1 + \beta\iint W(x_2-s_2(x_1,x_2))g(x_1,x_2,a_1,a_2(x_1))dx_1dx_2,$$

where $g(\cdot) = h(x_2|a_1,x_1,a_2(x_1))f(x_1|a_1)$.

The agent's expected utility is

$$\int U(s_1(x_1))f(x_1|a_1)dx_1 + \alpha\iint U(s_2(x_1,x_2))g(x_1,x_2,a_1,a_2(x_1))dx_1dx_2 - V(a_1)$$
$$- \alpha\int V(a_2(x_1))f(x_1|a_1)\, dx_1.$$

The interior portions of the optimal sharing rules are characterized by

$$\frac{W'(x_1-s_1(x_1))}{U'(s_1(x_1))} = \lambda + \mu_1 \frac{f_1(x_1|a_1)}{f(x_1|a_1)} \tag{3.7.3}$$

and

$$\frac{\beta}{\alpha}\frac{W'(x_2-s_2(x_1,x_2))}{U'(s_2(x_1,x_2))} = \lambda + \mu_1 \frac{g_1(x_1,x_2,a_1,a_2(x_1))}{g(x_1,x_2,a_1,a_2(x_1))}$$

$$+ \mu_2(x_1) \frac{g_2(x_1,x_2,a_1,a_2,(x_1))}{g(x_1,x_2,a_1,a_2,(x_1))}, \tag{3.7.4}$$

where the subscripts j on the distributions indicate partial derivatives with respect to a_j.

Equation (3.7.4) is similar to the characterization for single-period sharing rules in the multidimensional effort problem. Thus, the multidimensional effort results described earlier are useful in extending the theory to a certain class of finite-horizon multiperiod problems. Lambert (1981) has examined the model above under the assumption that effort in one period has no effect on the outcome in any other period, and also examined problems which occur when the principal is committed to a two-period contract, but the agent can leave the firm after the first period.

The second special case of interest is the case in which the principal's and the agent's expected utilities depend only on the total return over the entire time horizon. In this case, the principal's utility function is $W(x_1 + x_2 - s(x_1,x_2))$, and the agent's utility is $U(s(x_1,x_2), a_1,a_2(\cdot))$. This structure is also appropriate for the problem of sequential allocation of effort within one time period, where the time period is said to end when the agent receives his or her compensation. The sequential allocation aspect would arise because of the agent's opportunity to observe an outcome affected by the first effort choice before making any other effort choices. This situation is the focus of the next chapter.

CHAPTER 3 FOOTNOTES

1. Although there are technical problems connected with the use of the normal distribution, it is used here for illustrative purposes because it is the only distribution with a convenient representation for dependent random variables. Detailed calculations and results for the normal distribution appear in Appendix 2.

2. A modified version of this result holds when the principal is risk averse. Differentiating the first-order condition characterizing the sharing rule with respect to x shows that

$$\text{sign}\ (s'(x)) = \text{sign}\ \left(\mu \frac{\partial}{\partial x} \frac{f_a}{f} - \frac{W''U'}{U'^2}\right).$$

Thus, $\mu \frac{\partial}{\partial x} \frac{f_a}{f} \geq 0$ implies that $s'(x) \geq 0$, but the converse does not hold.

3. As Holmstrom (1979) points out, if the production function x is given by $x(a, \theta)$, where θ represents a random state of nature, then $\partial x/\partial a > 0$ implies that the distribution of x satisfies the first-order stochastic dominance property (provided that changes in a have a nontrivial effect on the distribution).

4. Extending this and the other propositions which depend on the assumption of a square root utility function to a more general class of utility functions appears to be nontrivial. However, in the discrete-outcome example presented earlier, the result is not confined to only the square root utility function. Hence, it appears likely that this and the other results stated for the square root utility function hold for a more general class of utility functions.

5. Lal (1982) also independently applied agency theory to the problem of sales force management. Much of his analysis is for a special normal distribution and the class of power utility functions. He did not analyze the additive effort case.

6. Let p be the constant sales price of a product, and c be the constant noncompensation cost per unit of product. Further, let q be the random quantity sold as a result of the agent's effort. One question of interest is whether the agent's compensation should be based on, for example, sales (pq) or a "contribution margin" (pq−cq). It is easy to see that the optimal sharing rule is characterized by $\frac{1}{U'(s(\cdot))} = \lambda + \mu \frac{f_a(q|a^*)}{f(q|a^*)}$. That is, the optimal sharing rule does not depend explicitly on p or c or p−c.

7. The formulation is presented in order to illustrate the structure of the problem. Technical problems with the properties of the functions to be maximized are not addressed.

CHAPTER 4

ONE-PERIOD SEQUENTIAL CHOICE

In this chapter, the model is expanded to include decisions made at different times. The extension is to sequential decisions within one period, where a period is defined to end at the time of payment to the agent. The one-period sequential case is an intermediate step between the allocation of effort case, in which both the efforts are exerted before the outcomes are known, and the two-period case, in which the first outcome is observed and the first compensation is paid before the second effort is exerted.

The allocation and sequential situations can be depicted as follows:

Allocation of effort:

Principal chooses $s(x_1,x_2)$	Agent exerts a_1,a_2	Principal and agent observe x_1,x_2; principal pays $s(x_1,x_2)$ to the agent.

One period sequential choices:

Principal chooses $s(x_1,x_2)$	Agent exerts a_1	Agent observes x_1	Agent exerts $a_2(\cdot)$	Agent observes x_2; principal observes x_1 and x_2 and pays $s(x_1,x_2)$ to the agent.

Two-period sequential choices:

Principal chooses $s_1(x_1)$ and $s_2(x_1,x_2)$	Agent exerts a_1	Principal and agent observe x_1; principal pays $s_1(x_1)$ the agent.	Agent exerts $a_2(\cdot)$	Principal and agent observe x_2; principal pays $s_2(x_1,x_2)$ to the agent.

In each of the cases above, if the principal and the agent observe additional valuable post-decision information about the agent's efforts, then the sharing rules will depend on this additional information.

A number of situations might be modeled in the one-period sequential framework. In the sales force management example, the agent might spend a certain amount of time selling products in one territory and observe the amount of the resultant sales there before beginning work in another territory. If there is correlation between x_1 and x_2, then the agent obtains information from x_1 which may be useful in the decision about a_2. The additional post-decision information that the principal may obtain about the agent's efforts might be comments obtained from personally interviewing the agent's customers.

Another one-period sequential decision setting might involve production decisions by an agent, where a_1 is the number of hours of production until some sales information is obtained. The agent would then choose the number of hours of production for the remainder of the period. In this situation, the additional post-decision information obtained by the principal might be the number of work hours recorded on the agent's time cards. More generally, a manager in a decentralized organization will not be monitored daily, but rather will make many decisions during a given time period and will be evaluated only periodically.

The one-period sequential model can be thought of as the special case of the fully general two-period model in which the periods are very short, so that the principal's and the agent's expected utilities depend only on their total return for the entire horizon. The one-period model can incorporate some of the elements of the fully general two-period model while providing a somewhat simplified structure for analysis. For example, in both models, the first outcome, which is first-stage post-decision information, can be used as pre-decision information for the second effort choice. The agent's precommitment to stay for the entire time horizon is not a major problem in the one-period model, since the agent is not paid until all the required efforts have been exerted.

In the first part of this chapter, the simplified structure in the one-period sequential model is used to explore the impact of correlation of outcomes in first best and second best situations. Some comparisons are made to the allocation of effort results. The analysis will focus on aspects which were not addressed in the pre-decision information literature or in Lambert's (1983) analysis of a finite-horizon multiperiod agency problem with independent outcomes. The second part of this chapter develops results for the one-period sequential problem that parallel two sets of results in the allocation of effort problem, namely additive separability of the sharing rule and diversification of effort across tasks when effort is additive. The similarities to and differences from the allocation results are discussed.

Before proceeding to the analysis, a brief review of the existing results on pre-decision information will be given and Lambert's (1983) results will be summarized. Unless otherwise stated, the "sequential effort problem" will refer to the one-period sequential effort problem.

A limited amount of research has been devoted to one-period agency problems with pre-decision information. Baiman (1982, p. 192) comments as follows on the increased complexity with pre-decision information:

> The role and value of a pre-decision information system is more complex than that of a post-decision information system. Expanding a post-decision information system to report an additional piece of information will always result in at least a weak Pareto improvement, since the principal and agent can always agree to a payment schedule that ignores the additional information. However, expanding a pre-decision information system to report an additional piece of information may not result in even a weak Pareto improvement. The agent generally cannot commit himself to ignore the additional information, and therefore the optimal employment contract without the additional pre-decision information is no longer necessarily self-enforcing given the additional information. This is true whether the additional pre-decision information is privately reported or publicly reported.

Some of the research concerning pre-decision information focuses on the following question: Given that the agent has private pre-decision information, what is the value of public post-decision information systems? Holmstrom (1979) showed that an informativeness criterion ($f(x,y,z;a) \neq g(x,y)h(x,z;a)$, where z is the pre-decision signal) is necessary for the post-decision information system which reports a public signal, y, in addition to x, the outcome, to provide a Pareto improvement over the information system which reports only x. Christensen (1982) expanded Holmstrom's (1979) model by allowing the agent to communicate to the principal a message m about the private pre-decision signal. The agent is assumed to select the message that maximizes his or her expected utility. In Christensen's model, a generalization of Holmstrom's (1979) informativeness criterion is necessary for the post-decision information system which reports y, in addition to x and m, to provide a Pareto improvement over the information system which reports only x and m. Here, the public post-decision signal is a signal about the agent's effort and the agent's private pre-decision information signal.

Another direction of the research on pre-decision information has been the value of pre-decision information systems. There are both positive and negative effects of private pre-decision information for the agent. On one hand, the agent has more information before choosing an action, and hence should make "better" decisions. On the other hand, more information may reduce the risk the agent faces, and hence reduce the motivation for the agent to exert effort. Christensen (1981) provided an example which shows that the principal may be worse off when the agent has private pre-decision information (with or without communication of a message), and also provided an example which shows that the principal may be better off when the agent has private pre-decision information and communicates a message to the principal. Christensen's examples illustrate the difficulty in obtaining a general preference ordering rule over pre-decision information systems.

A third direction of research on private pre-decision information has been the value of communication of a message about the private information from the agent to the principal, given the existence of the private pre-decision information system. In the accounting context, the focus is on the value of communication of private information in the process of participative budgeting. The major result in this area is that of Baiman and Evans (1983), who provided necessary and sufficient conditions for communication to result in a Pareto improvement. Baiman (1982, p. 204) summarizes the result as follows:

> . . . If the agent's private pre-decision information is perfect, then communication has no value. Observing the firm's output in that case allows the principal to infer all he needs to know about the agent's private pre-decision information. However, if the agent's private pre-decision information is imperfect, a necessary and sufficient condition for communication to be strictly valuable is for the honest revelation of the agent's private pre-decision information to be strictly valuable. That is, if any value can be achieved with the information being honestly revealed to <u>all</u>, then a strictly positive part of that value can be achieved by giving the agent sole direct access to the information and letting him communicate in a manner that maximizes <u>his</u> expected utility.

Lambert (1983) has examined a special case of the finite-horizon multiperiod agency problem. He assumed that both the principal and the agent have utility functions (and that the agent has a disutility function) which are separable across time. He further assumed that the state variables are independently distributed across time, and that effort in one period does not influence the monetary outcome in any other period. Under these conditions, Lambert showed that the agent's compensation in a given period will depend on the outcomes in previous periods as well as on the outcome in the present period. He further showed that the incentive problems associated with the agent's effort choices in each period are not eliminated. In the notation of this chapter, the result can be stated as (i) $\mu_1 > 0$, and (ii) $\mu_2(x_1) > 0$ for almost every x_1 (first-stage outcome).

The remainder of this chapter analyzes the one-period sequential effort choice problem. The cooperative, or first best case is first considered, and the behavior of the agent's second-stage effort choice strategy is characterized. The second best case is then analyzed. The optimal sharing rule is derived and discussed, as is the behavior of the agent's second-stage choice strategy, with and without independence of the outcomes. It is then shown that the optimal sharing rule will not be additively separable in the outcomes, even under the conditions which were sufficient for such a result in the effort allocation problem. Finally, the special case of additive effort is analyzed, and the question of the desirability of diversification of the agent's efforts across tasks is examined. The result is related to the information content of the outcome about the agent's effort.

4.1. FIRST BEST

In the first best case, the principal's problem is:

$$\underset{s(\cdot),a_1,a_2(\cdot)}{\text{Maximize}} \quad \int\int W(x-s(x_1,x_2))\phi(x_1,x_2|a_1,a_2(\cdot))\,dx_2dx_1$$

$$\text{subject to} \int\int \{U(s(x_1,x_2)) - V(a_1,a_2(\cdot))\}\phi(\cdot)dx_2dx_1 \geq \bar{u},$$

where $\phi(x_1,x_2|a_1,a_2(\cdot)) = f(x_1|a_1)g(x_2|x_1,a_1,a_2(\cdot))$ and $a_2(\cdot)$ indicates that the agent's second-stage effort is in general not a constant, but rather can depend on any information available at the time of choice. Letting λ be the multiplier for the agent's expected utility constraint and differentiating the Hamiltonian with respect to $s(\cdot)$ for every (x_1,x_2) yields

$$\frac{W'(x-s(x_1,x_2))}{U'(s(x_1,x_2))} = \lambda$$

for almost every (x_1,x_2). This implies that if one person is risk neutral and the other is risk averse, then the risk neutral person will bear the risk (see Appendix 4). That is, if the principal is risk neutral and the agent is risk averse, then the optimal sharing rule is constant; if the principal is risk averse and the agent is risk neutral, then the principal's return is k, a constant, and the agent receives x_1+x_2-k. If both individuals are risk averse, then the risk is shared; the optimal sharing rule is a function of (x_1+x_2). Furthermore, $\partial s/\partial x_i$ is positive for $i = 1,2$. Finally, if both are risk neutral, then the optimal sharing rule is $s = \bar{u} + V(a_1,a_2(\cdot))$. These results are the same as those for the allocation of effort problem. Thus, in the first best case, the sequential nature of the effort decisions does not affect the characterization of the optimal sharing rules.

In this scenario, there are no signals on which the choice of a_1 can be based. Whether or not a_2 is a function of x_1 depends on the risk attitudes of the individuals and the joint distribution $\phi(x_1,x_2|a_1,a_2(\cdot))$. If at least one of the individuals is risk neutral and $\phi(x_1,x_2|a_1,a_2(\cdot)) = f(x_1|a_1)g(x_2|a_2(\cdot))$, then the optimal $a_2(\cdot)$ is independent of x_1. In this case, the risk neutral person essentially owns the output of the firm, and thus bears all the risk associated with the uncertainty of x_1. Furthermore, x_1 conveys no information about x_2.

If both of the individuals are risk averse or if $\phi(\cdot)$ is the more general $f(x_1|a_1)g(x_2|x_1,a_1,a_2(\cdot))$, then the optimal $a_2(\cdot)$ will generally depend on x_1. In the first case, the change from the situation where one individual is risk neutral occurs because each risk averse individual's marginal utility depends on the first outcome, since it determines where on his or her utility curve the individual is; a risk neutral individual's marginal utility, on the other hand, would be the same no matter what the value of x_1 is. This first effect of x_1 can be termed the "wealth" or "risk aversion" effect. In the second case, if x_1 and x_2 are dependent, then expectations for x_2 may change according to the first outcome, x_1. The principal may therefore wish to induce the agent to choose $a_2(\cdot)$ as an increasing or decreasing function of x_1, depending on the risk attitudes of the principal and the agent, the agent's disutility for effort, and the nature of the correlation between x_1 and x_2. This second effect of x_1 can be termed the "information" effect. The information effect of x_1 is made more precise in the proposition below.

Proposition 4.1.1. Suppose that in the first best case, the principal is risk neutral, the agent is risk averse, and $\phi(\cdot) = f(x_1|a_1)g(x_2|x_1,a_1,a_2(\cdot))$. In this case, $a_2(\cdot)$ will depend on x_1. Let $M_1(\cdot)$ denote the mean of x_1 given a_1, and let $M_2(x_1,a_1,a_2(\cdot))$ denote the conditional mean of x_2 with respect to $g(\cdot)$. Let the second and third subscripts of j on M_2 denote partial differentiation of M_2 with respect to the j-th argument of $M_2(x_1,a_1,a_2(\cdot))$. Then

$$a_2^{*\prime}(x_1) = -M_{231}/[M_{233} - \lambda[\partial^2 V(\cdot)/\partial a_2^2]].$$

For example, suppose $M_2(x_1,a_1,a_2(\cdot)) = x_1 a_2(x_1)/a_1$ and $V(\cdot) = (a_1+a_2)^2$. Then $a_2^{*\prime}(x_1) = 1/(2a_1^*\lambda) > 0$. In this case, $a_2^*(x_1)$ increases linearly in x_1.

The effect of the nature of the correlation between x_1 and x_2 is captured in the derivatives of $M_2(\cdot)$, and the effect of the disutility function is captured in the $\partial^2 V/\partial a_2^2$ term. Note that $a_2^*(x_1)$ does not depend on the agent's utility function for wealth. This is because the risk averse agent receives a constant wage in the first best case, and hence the agent's utility for the wage is constant. Note further that if M_2 depends only on $a_2(\cdot)$, then a_2^* is constant.

Proposition 4.1.1 and the discussion preceding it focused on the second effort choice's dependence on x_1, the first outcome. The second effort choice, $a_2(\cdot)$, might seem to also depend on the first effort choice, a_1.

However, the agent chooses the effort a_1 and the effort strategy $a_2(\cdot)$ simultaneously at the beginning of the time horizon. The second effort choice is therefore not viewed as a function of a_1, although there is implicit recognition that a_1 and $a_2(\cdot)$ are chosen jointly and therefore influence one another. However, since the first outcome is unknown at the beginning of the time horizon, the second effort choice can potentially depend on the first outcome.

4.2 SECOND BEST

In this section, the general formulation of the one-period sequential model is first presented. Subsequently, the two extremes of independent outcomes and perfectly correlated outcomes are examined. In the first case, knowledge of x_1 reveals no information about x_2, whereas in the second case, x_1 reveals perfect information about x_2. The behavior of the agent's second effort strategy is illustrated in the two extreme cases, and also for the intermediate case of imperfectly correlated outcomes.

As before, in order to focus on motivational issues, it will be assumed that the principal is risk neutral and the agent is risk averse. The principal's problem is:

$$\underset{s(\cdot),a_1,a_2(\cdot)}{\text{Maximize}} \quad \int\int (x-s(x_1,x_2))\,\phi(x_1,x_2|a_1,a_2(\cdot))dx_2dx_1$$

subject to

$$\int [\,\int U(s(x_1,x_2))g(x_2|x_1,a_1,a_2(\cdot))dx_2 - V(a_1,a_2(\cdot))]\,f(x_1|a_1)dx_1 \geq \bar{u}$$

$$\int\int U(s(\cdot))[g_{a_1}f + gf_{a_1}]dx_2dx_1 - \int (V_{a_1}f + Vf_{a_1})dx_1 = 0$$

$$\{\int U(s(\cdot))g_{a_2}(\cdot)dx_2 - V_{a_2}(\cdot)\}f(x_1|a_1) = 0 \text{ for almost every } x_1,$$

where $\phi(x_1,x_2|a_1,a_2(\cdot)) = f(x_1|a_1)g(x_2|x_1,a_1,a_2(\cdot))$ and differentiation with respect to a_2 is pointwise for each x_1. The interior portion of the optimal sharing rule is characterized by

$$\frac{1}{U'(s(\underline{x}))} = \lambda + \mu_1 \phi_{a_1}/\phi + \mu_2(x_1)\phi_{a_2}/\phi \text{ for almost every } (x_1,x_2),$$

where λ, μ_1 and $\mu_2(x_1)$ are multipliers for the three constraints above. Here, $\phi_{a_1}/\phi = f_{a_1}/f + g_{a_1}/g$ and $\phi_{a_2}/\phi = g_{a_2}/g$. If a_1 does not influence x_2, then $\phi_{a_1}/\phi = f_{a_1}/f$.

The characterization of the interior portion of the optimal sharing rule in the sequential effort case is similar to that in the allocation of effort case, except that here μ_2^* and a_2^* may depend on x_1. In general, $a_2^*(\cdot)$ depends on x_1. However, if the agent is risk neutral and $\phi(x_1,x_2|a_1,a_2(\cdot)) = f(x_1|a_1)g(x_2|a_2(\cdot))$, then $a_2^*(\cdot)$ does <u>not</u> depend on x_1. If the x_i's are conditionally correlated, then $a_2^*(\cdot)$ will depend on x_1 even if the agent is risk neutral. These results are direct consequences of the achievability of the first best solution in the second best case if the agent is risk neutral (see Shavell (1979)).

The proposition below describes aspects of the second stage problem for a particular utility function for the agent, and for several commonly used distributions for the independent outcomes.

<u>Proposition 4.2.1</u>. Suppose that in the second best case, the principal is risk neutral, and the agent's utility function for wealth is $U(s) = 2\sqrt{s}$. Suppose also that $\phi(\cdot) = f(x_1|a_1)g(x_2|a_2(\cdot))$, where $f(\cdot)$ and $g(\cdot)$ are in Q_1, the class consisting of the exponential, gamma, and Poisson distributions represented in Appendix 1. Define a_1 and a_2 so that the mean of $f(x_1|a_1)$ is a_1 and the mean of $g(x_2|a_2)$ is a_2. Then, assuming that the optimal efforts are nonzero,

(i) if $\partial V/\partial a_2$ is positive at $\underline{a}^*$ then

(a) $\mu_2(x_1)$ is positive, and

(b) a sufficient condition for the agent's expected second stage net utility to be increasing in x_1 is that $a_2^*(\cdot)$ be a decreasing function of x_1;

(ii) if μ_1 is positive, then

(a) the agent's expected utility for the second stage pecuniary return, $E\{U(s(\underline{x}))|x_1\}$, is an increasing function of x_1, and

(b) the conditions $V_2 > 0$, $V_{22} \geq 0$, $V_{222} \geq 0$, and $V_{122} \geq 0$ are jointly sufficient for $a_2^*(\cdot)$ to be a decreasing function of x_1. Here, subscripts j on V represent partial differentiation with respect to the j-th effort variable.

The condition that $\partial V/\partial a_2$ be positive is a standard one, and is nonrestrictive. A number of general forms of disutility functions satisfy the conditions in (ii)(b). The following, for example, satisfy the conditions:

$V(a_1,a_2) = h(a_1) + a_2^m$, where $m \geq 1$ and $a_2 > 0$,

$V(a_1,a_2) = a_1^2 a_2^2$ where $a_1 \geq 0$ and $a_2 \geq 0$, and

$V(a_1,a_2) = h(c_1a_1 + c_2a_2)$, where $h' > 0$, $h'' \geq 0$, $h''' \geq 0$, and the constants c_1 and c_2 are positive.

If μ_1 is zero, so that there is no incentive problem, then a_2 does not depend on x_1 (see Appendix 4). This is consistent with the first best results with a risk neutral principal, a risk averse agent, and independence of the outcomes. There is neither an incentive problem nor an information effect to induce the dependence of a_2 on x_1.

In general, though, a_2 will depend on x_1. Proposition 4.2.1 states that in some particular settings, the optimal second stage effort will decrease as the first outcome increases. Recall that x_1 determines a point on the utility curve for the agent before the second stage effort is chosen. Because the agent's marginal utility for wealth is a decreasing function and the agent's marginal disutility for effort is an increasing function, it is more costly for the principal to induce a given level of a_2, the higher x_1 is. The result that a_2 is decreasing in x_1 should thus hold for other concave utility functions for wealth, coupled with convex disutility functions.

Proposition 4.2.1 also provides conditions under which the agent's second stage expected utility will increase as the outcome increases. Under the given conditions, $E[U(s(\underline{x}))|x_1]$ is increasing in x_1, and $-V(a_1,a_2(x_1))$ is increasing in x_1 because a_2 is decreasing in x_1. Thus, the agent's expected second stage net utility is increasing in x_1.

The independence of x_1 and x_2 in Proposition 4.2.1 means that there is no information effect of x_1. If x_1 and x_2 are correlated, then the behavior of $a_2^*(\cdot)$ would depend additionally on the nature of the correlation. In order to examine the information effect of x_1, the extreme case of perfect correlation of the outcomes will next be analyzed. When the outcomes are perfectly correlated, then a joint density for x_1 and x_2 does not exist. Since the lack of a joint density precludes using the previous analysis directly, a modified approach must be taken in order to examine the nature of the sharing rule and the agent's second-stage effort choice when the outcomes are perfectly correlated.

Let $x_i = x_i(\theta,a_i)$, where θ is an uncertain state that influences both the outcomes. It will be assumed that for any fixed a_1, x_1 can be inverted to obtain $\theta = \theta(x_1,a_1)$, The principal's and the agent's common beliefs about the outcomes will be expressed as $\phi(x_1,x_2|a_1,a_2(\cdot)) = f(x_1|a_1)$ if $x_2 = x_2(\theta,a_2(x_1))$ and $\theta = \theta(x_1,a_1)$; otherwise, $\phi(\cdot) = 0$.

In order to describe the sharing rule, let a_1^* be the agent's first-stage effort choice that is induced by the sharing rule, and let $a_2^*(x_1)$ be the agent's second-stage effort strategy that is induced if x_1 is observed and it is assumed that $a_1 = a_1^*$. Because of the perfect correlation between

x_1 and x_2, the sharing rule $\hat{s}(x_1,x_2)$ can be viewed as being of the following form: $\hat{s}(x_1,x_2) = s(x_1)$ if $x_2 = x_2(\theta,a_2^*(x_1))$ and $\theta = \theta(x_1,a_1^*)$; otherwise, $\hat{s}(\cdot)$ is a penalty wage which is possibly negative.

The sharing rule can be viewed as being dichotomous with respect to x_2 and varying continuously only with x_1. Alternatively, the sharing rule can be viewed as being a function of the total output, $x_1 + x_2$, subject to the condition that the observed x_2 is in agreement with the observed value of x_1 and the inferred value of θ. In either view of the sharing rule, lack of agreement between the observed values of x_2 and x_1 is taken as evidence of shirking; accordingly, a penalty is imposed in such situations. If the penalty is sufficiently severe, the penalty need never be imposed, since the agent will choose to avoid the penalty by choosing $a_2^*(x_1)$. Determination of the optimal sharing rule can hence be confined to determination of the optimal function $s(x_1)$; furthermore, no first order condition is required in order to induce $a_2^*(x_1)$, as long as a_1^* is properly induced.

The principal's problem can therefore be written as follows:

$$\underset{s(x_1),a_1,a_2(x_1)}{\text{Maximize}} \quad \int W(x_1 + x_2(\theta(x_1,a_1),a_2(x_1)) - s(x_1))f(x_1|a_1)dx_1$$

$$\text{subject to} \quad \int[U(s(x_1)) - V(a_1,a_2(x_1))]f(x_1|a_1)dx_1 \geq \bar{u}$$

$$\int[U(s(x_1)) - V(a_1,a_2(x_1))]f_{a_1}(x_1|a_1)dx_1 - \int V_{a_1}(a_1,a_2(x_1))f(x_1|a_1)dx_1 = 0.$$

In order to determine the first order conditions, let λ and μ be the Lagrangian multipliers for the first and second constraints, respectively, and form the Hamiltonian H in the usual way. Differentiating H with respect to $s(\cdot)$ for every x_1 yields

$$\frac{W'(x_1 + x_2(\theta(x_1,a_1),a_2(x_1)) - s(x_1))}{U'(s(x_1))} = \lambda + \mu \frac{f_{a_1}(x_1|a_1)}{f(x_1|a_1)},$$

which is of the usual form. Differentiating H with respect to a_1 yields

$$\int W'(\cdot)\frac{\partial x_2}{\partial\theta}\frac{\partial\Theta}{\partial a_1}f(\cdot)dx_1 + \int W(\cdot)f_{a_1}(\cdot)dx_1 + \mu\int\{[U(\cdot) - V(\cdot)]f_{a_1a_1}(\cdot)$$

$$- 2V_{a_1}(\cdot)f_{a_1}(\cdot) - V_{a_1a_1}(\cdot)f(\cdot)\}dx_1 = 0.$$

Finally, differentiating H with respect to a_2 for every x_1 yields

$$W'(\cdot)\frac{\partial x_2}{\partial a_2}f(\cdot) - \lambda V_{a_2}(\cdot)f(\cdot) - \mu[V_{a_2}(\cdot)f_{a_1}(\cdot) + V_{a_1a_2}(\cdot)f(\cdot)] = 0.$$

If the principal is risk neutral, as is commonly assumed in order to focus on motivational rather than risk-sharing issues, then the first order conditions above reduce to

$$\frac{1}{U'(s(x_1))} = \lambda + \mu\frac{f_{a_1}(x_1|a_1)}{f(x_1|a_1)}, \tag{4.2.1}$$

$$\int\frac{\partial x_2}{\partial\theta}\frac{\partial\theta}{\partial a_1}f(x_1|a_1)dx_1 + \int W(\cdot)f_{a_1}(x_1|a_1)dx_1$$

$$+ \mu\int\{[U(s(x_1)) - V(a_1,a_2(x_1))]f_{a_1a_1}(x_1|a_1)dx_1$$

$$- 2V_{a_1}(a_1,a_2(x_1))f_{a_1}(x_1|a_1) - V_{a_1a_1}(a_1,a_2(x_1))f(x_1|a_1)\}dx_1 = 0, \tag{4.2.2}$$

and

$$\frac{\partial x_2}{\partial a_2}f(x_1|a_1) - \lambda V_{a_2}(a_1,a_2(x_1))f(x_1|a_1)$$

$$- \mu[V_{a_2}(a_1,a_2(x_1))f_{a_1}(x_1|a_1) + V_{a_1a_2}(a_1,a_2(x_1))f(x_1|a_1)] = 0. \tag{4.2.3}$$

Dividing (4.2.3) by $f(x_1|a_1)$ and rearranging yields

$$\frac{\partial x_2}{\partial a_2} + [\lambda + \mu\frac{f_{a_1}(x_1|a_1)}{f(x_1|a_1)}]V_{a_2}(a_1,a_2(x_1)) + \mu V_{a_1a_2}(a_1,a_2(x_1)). \tag{4.2.4}$$

Substituting (4.2.1) into (4.2.4) yields

$$\frac{\partial x_2}{\partial a_2} = \frac{1}{U'(s(x_1))}V_{a_2}(a_1,a_2(x_1)) + \mu V_{a_1a_2}(a_1,a_2(x_1)). \tag{4.2.5}$$

It is easily seen that if (4.2.4) is to hold for almost every x_1, then $a_2(\cdot)$ must in general vary with x_1. The "wealth" and "information" effects of x_1 described in the first best analysis can be seen in (4.2.4). The wealth effect of x_1 results from the interaction of the agent's marginal utility for wealth and marginal disutility for effort. The information effect of x_1 refers to the information that x_1 provides about x_2. In the perfect correlation case, the state θ is inferred from x_1 and a_1^*, and x_2 is hence a deterministic function of a_2 from both the principal's and the agent's perspectives. The information effect of x_1 is therefore captured in the $\partial x_2/\partial a_2$ term in (4.2.5).

The behavior of a_2^* as x_1 varies can be determined by differentiating (4.2.4) with respect to x_1 to obtain

$$a_2^{*\prime}(x_1) = \frac{\mu V_{a_2} \frac{\partial}{\partial x_1}\left(\frac{f_{a_1}}{f}\right) - \frac{\partial}{\partial \theta}\left(\frac{\partial x_2}{\partial a_2}\right)\frac{\partial \theta}{\partial x_1}}{\frac{\partial^2 x_2}{\partial a_2^2} - \left(\lambda + \mu \frac{f_{a_1}}{f}\right)V_{a_2 a_2} - \mu V_{a_1 a_2 a_2}} . \tag{4.2.6}$$

When x_2 is linear in a_2, the $\partial^2 x_2/\partial a_2^2$ term in the denominator is zero. Two special cases of interest are (i) $x_1 = \theta + a_1$, where θ is purely noise, and (ii) $x_1 = \theta a_1$, where θ reveals information about the production technology. In case (i), the marginal output per unit of effort is one, regardless of the value of θ. In case (ii), however, the marginal output per unit of effort is θ.

To illustrate the results, suppose that $V(\cdot) = a_1^2 a_2^2$. For case (i), assume that $\theta \sim N(0,\sigma^2)$. Then $x_1 \sim N(a_1,\sigma^2)$ and $f_{a_1}/f = (x_1 - a_1)\sigma^2$. Therefore, the numerator of (4.2.6) is $\mu(2a_1^2 a_2)/\sigma^2$ and the denominator is $-(\lambda + \mu \frac{x_1 - a_1}{\sigma^2})(2a_1^2) - 4a_1\mu$. The term $(\lambda + \mu \frac{x_1 - a_1}{\sigma^2})$ is positive by the first order condition (4.2.1), and the effort levels are assumed to be positive. Therefore, $a_2^{*\prime}(x_1) < 0$, provided that $\mu > 0$. This can also be seen by solving for $a_2(x_1)$ directly from (4.2.4) to obtain

$$a_2(x_1) = [2a_1^2(\lambda + \mu \frac{x_1-a_1}{\sigma^2}) + 4\mu a_1]^{-1}.$$

In this case, the sign of $a_2^{*\prime}$ is the same as in the independent outcome situation described in Proposition 4.2.1, where there was no information about x_2 to be gained from x_1. The case (i) result here can thus be interpreted as indicating that the wealth effect of x_1 dominates any information effect that exists through perfect correlation of the outcomes.

For case (ii), assume first that $\theta \sim \exp(1)$. Then $x_1 \sim \exp(a_1)$ and $f_{a_1}/f = (x_1-a_1)/a_1^2$. Equation (4.2.4) becomes

$$\theta = (\lambda + \mu \frac{x_1-a_1}{a_1^2})2a_1^2 a_2 + 4\mu a_1 a_2.$$

Substituting $\theta = x_1/a_1$ and rearranging results in

$$a_2^*(x_1) = \frac{x_1}{2a_1^2[a_1(\lambda + \mu \frac{x_1-a_1}{a_1^2}) + 2\mu]}.$$

Therefore,

$$a_2^{*\prime}(x_1) = \frac{1}{2a_1^2}\left[\frac{a_1(\lambda + \mu \frac{x_1-a_1}{a_1^2}) + 2\mu - x_1(\mu/a_1)}{\{a_1(\lambda + \mu \frac{x_1-a_1}{a_1^2}) + 2\mu\}^2}\right].$$

The numerator of $a_2^{*\prime}(x_1)$ reduces to $(a_1\lambda - \mu + 2\mu)$, which is positive (assuming $\mu > 0$) because $a_1(\lambda + \mu \frac{x_1-a_1}{a_1^2}) > 0$ for $x_1 \geq 0$, and for $x_1 = 0$ in particular. Thus, $a_2^{*\prime}(x_1) > 0$ in this case. A similar analysis can be done for the normal distribution example used in case (i), with the result that

$a_2^{*\prime}(x_1) > 0$. The sign of $a_2^{*\prime}(x_1)$ would remain the same in cases (i) and (ii) for a wide variety of reasonable disutility functions.

For the normal distribution example, the only difference in the expression for $a_2^{*\prime}(x_1)$ is the $\frac{\partial}{\partial\theta}\left(\frac{\partial x_2}{\partial a_2}\right)\frac{\partial\theta}{\partial x_1}$ term. In case (i), it is zero, and in case (ii), it is $1/a_1$. Although θ is purely noise in case (i), risk is imposed on the agent for motivational purposes in order to induce a Pareto optimal choice of a_1. The effort strategy $a_2(x_1)$ is primarily determined by the wealth effect of x_1, leading to a decreasing function of x_1 just as in the case when the outcomes were assumed to be independent (see Proposition 4.2.1). In case (ii), where the marginal output per unit of effort is θ, the agent receives perfect information about the production technology that was not relevant in case (i). The information effect of x_1 overrides the wealth effect in the case (ii) examples above, so that a_2^* is now an increasing function of x_1.

To illustrate the second best results, suppose that the principal is risk neutral and the agent's utility for wealth is $2\sqrt{s}$. Suppose further that $x_1|a_1$ and $x_2|a_2$ are independent, $f(\cdot)$ is exponential with mean a_1, and $g(\cdot)$ is exponential with mean $a_2(x_1)$. Then the interior portion of the optimal sharing rule is characterized by $s(x_1,x_2) = P^2(\underline{x})$, where

$$P(\underline{x}) = \lambda + \mu_1\left(\frac{x_1-a_1^*}{a_1^{*2}}\right) + \mu_2(x_1)\left(\frac{x_2-a_2^*(x_1)}{a_2^{*2}(x_1)}\right). \tag{4.2.7}$$

$P(\underline{x})$ must be strictly positive in order to satisfy the first order condition $1/U' = P(\underline{x})$. In the proof of Proposition 4.2.1, it is shown that

$$\mu_2(x_1) = (\partial V(\underline{a}^*)/\partial a_2)a_2^{*2}(x_1)/2 , \tag{4.2.8}$$

which is positive under the usual assumption that the agent's disutility function is increasing in the second effort. Furthermore, it is easily seen that $\mu_2'(x_1) < 0$ under the assumptions in Proposition 4.2.1, part (ii). Intuitively, the higher the first outcome is, the less concerned the risk neutral principal is about motivating a high choice of a_2. This is because the higher the outcome x_1 is, the costlier it becomes to induce a given level of a_2. As remarked earlier, the principal induces a strategy $a_2(x_1)$ which is decreasing in x_1.

At the time of the second effort choice, the first outcome x_1 is known. As Lambert (1983) notes, $P(\underline{x})$ can be viewed as $\lambda(x_1) + \mu_2(x_1)(\frac{x_2 - a_2^*(x_1)}{a_2^{*2}(x_1)})$, which is as it would appear in a one-stage, one-period agency problem, given that x_1 is fixed. Thus, it is not totally surprising that, as in the one-stage, one-period problem, $\partial s/\partial x_2$ is strictly positive, since $P(\underline{x})$ and $\mu_2(x_1)$ are strictly postive. The behavior of $s(\cdot)$ as x_1 varies is considerably more complicated. Substituting (4.2.8) into (4.2.7) and differentiating shows that

$$\frac{\partial s}{\partial x_1} = 2P(\underline{x})[\frac{\mu_1}{a_1^{*2}} + (\frac{x_2 - a_2^*(\cdot)}{2})a_2^{*\prime}V_{a_2a_2}(\cdot) - a_2^{*\prime}V_{a_2}(\cdot)/2].$$

Under the assumptions in Proposition 4.2.1, part (ii), the first and third terms in the brackets are positive. The condition that $x_2 < a_2^*(x_1)$ is sufficient for the sharing rule to be increasing in the first outcome. However, it is clearly possible that $\partial s/\partial x_1$ is increasing in x_1 even if $x_2 > a_2^*(x_1)$.

An alternative approach to an analysis of the sharing rule is insightful. Recall that the first order conditions require that

$$\frac{1}{U'(s(\underline{x}))} = R(\underline{x}) = \lambda + \mu_1 \frac{f_{a_1}(x_1|a_1^*)}{f(x_1|a_1^*)} + \mu_2(x_1) \frac{g_{a_2}(x_2|a_2^*(x_1))}{g(x_2|a_2^*(x_1))} .$$

Taking the conditional expectation of $R(\underline{x})$ with respect to $g(x_2|a_2^*(x_1))$ results in the expression $\lambda + \mu_1 \frac{f_{a_1}(\cdot)}{f(\cdot)}$. As in the one-stage, one-period model, if $\mu_1 > 0$ annd $f(\cdot)$ satisfies the monotone likelihood ratio property, then $(\lambda + \mu_1 \frac{f_{a_1}(\cdot)}{f(\cdot)})$ is increasing in x_1. Thus, the agent faces a sharing rule with similar characterization for each stage, looking only one step ahead. That is, at each stage, the sha#ing rule is characterized by the condition that $1/U' = \lambda_0 + \mu_0 h_a/h$.

In order to illustrate the behavior of $a_2(x_1)$ when $x_1|a_1$ and $x_2|a_2$ are imperfectly correlated, suppose that $g(x_2|x_1,a_2(x_1))$ is exponential with mean $M_2(x_1) = x_1^b a_2(x_1)$. Since the exponential distribution is a one-parameter distribution, we may write $g(x_2|x_1,a_2(x_1)) = g(x_2|M_2(x_1))$, and Proposition 4.2.1 can be applied. For concreteness, suppose that $V(a_1,a_2) = a_1^2 a_2^2$. Then $V_2 = 2a_1^2 a_2 > 0$, $V_{22} = 2a_1^2 > 0$, $V_{222} = 0$, and $V_{122} = 4a_1 > 0$, so that the conditions in Proposition 4.2.1, part (ii)(b) are satisfied. Substituting $a_2(x_1) = M_2(x_1)/x_1^b$ into the expression for V_2 yields $V_2 = 2a_1^2 M_2(x_1)/x_1^b$, which is still positive. Therefore, if μ_1 is positive, then $M_2(x_1)$ is decreasing in x_1. That is, $x_1^b a_2(x_1)$ is decreasing in x_1. If b is positive, then it is easily seen that $a_2(x_1)$ is decreasing in x_1, as when b is zero (the "independent" case). In this situation, as when there is perfect correlation with the normal distribution in case (i), the wealth effect of x_1 is dominant. Recall that case (ii) of the perfect correlation analysis assumed that $x_1 = \theta a_1$, so that $x_2 = x_1 a_2(x_1)/a_1$. This seems similar to the imperfect correlation example in which $M_2(x_1) = x_1 a_2(x_1)$. However, the signs of $a_2^{*\prime}(x_1)$ are opposite in these perfect and imperfect correlation cases. This can be interpreted as follows: in the presence of information related to the production technology, the wealth effect of x_1 is dominant if the correlation is imperfect; the information effect of x_1 is dominant only if the correlation is perfect.

If b is negative, then the behavior of $a_2(x_1)$ is potentially much more complex. The condition that $M_2(x_1)$ is decreasing in x_1 is equivalent to the condition that $bx_1^{b-1}a_2(x_1) + x_1^b a_2'(x_1) < 0$. Since $b < 0$, the first term is negative; $a_2'(x_1)$ may thus be of any sign. It could be, for example, that because of the interactions of the wealth and information effects of x_1, $a_2(x_1)$ is increasing for low values of x_1 and decreasing for high values of x_1.

This concludes the analysis of the effect of the information x_1 on the agent's second effort strategy. The next two sections examine two aspects which were of interest in the allocation problem, namely additive separability of the sharing rule, and additive effort.

4.3. ADDITIVE SEPARABILITY OF THE SHARING RULE

In this section, the question of whether or not to reward the agent for each outcome separately is examined. For example, suppose a salesperson exerts effort selling a product in one territory, observes the resultant sales, and then devotes effort to selling the same product or a different product in another territory. Should the firm compensate the salesperson with a different reward function for each outcome, as if he or she were two separate salespeople? That is, should the sharing rule be additively separable in the outcomes?

It was shown in Section 3.4 that in the effort allocation problem, if the principal is risk neutral and the agent is risk averse with a HARA-class utility for wealth, then jointly sufficient conditions for the optimal sharing rule to be additively separable in x_1 and x_2 are (i) the agent has a log utility function for wealth and (ii) the outcomes are conditionally independent (see equation (3.3.2)). In the one-period sequential effort problem, the optimal sharing rule will not be additively separable in x_1 and x_2, even under conditions (i) and (ii) above. This is easily seen from the characterization of the interior portion of the optimal sharing rule:

$$s(x_1,x_2) = \begin{cases} \frac{1}{C}[(\nabla(\underline{x}))^C - D_2] & \text{if } C \neq 0 \\ D_2 \ln \nabla(\underline{x}) & \text{if } C = 0, \end{cases}$$

where the agent's risk aversion function is $-U''(s)/U'(s) = 1/(Cs+D_2)$ and

$$\nabla(\underline{x}) = \lambda + \mu_1 \frac{\phi_{a_1}(\cdot)}{\phi(\cdot)} + \mu_2(x_1) \frac{\phi_{a_2}(\cdot)}{\phi(\cdot)}$$

$$= \lambda + \mu_1 [\frac{f_{a_1}(x_1|a_1)}{f(x_1|a_1)} + \frac{g_{a_1}(x_2|x_1,a_1,a_2(\cdot))}{g(x_2|x_1,a_1,a_2(\cdot))}] + \mu_2(x_1) \frac{g_{a_2}(\cdot)}{g(\cdot)},$$

and differentiation with respect to a_2 is pointwise for every x_1. Thus, even if $U(s) = \ln s$ (i.e., $C = 1$) and $g(x_2|x_1,a_1,a_2(x_1)) = g(x_2|a_2(x_1))$, the optimal sharing rule will not be additively separable in x_1 and x_2 because of the $\mu_2(x_1)g_{a_2}/g$ term unless $\mu_2(x_1) \equiv k$, a constant, and g_{a_2}/g is additively separable in x_1 and x_2. Lambert (1981, p. 90) has shown in a similar situation that $\mu_2(x_1) > 0$ for almost every x_1. Since for almost every x_1, $\mu_2(x_1) \neq 0$, and it is unlikely that $\mu_2(x_1) \equiv k$ (which would require that $\partial E(x-s(\cdot))/\partial a_2 \equiv k\partial^2 E(U(s(\cdot)) - V(\cdot))/\partial a_2^2$ for almost every x_1), the optimal sharing rule will almost certainly not be additively separable in x_1 and x_2.

A corollary of this result is that if the principal is risk neutral and the agent's utility for wealth is in the HARA class, then the optimal sharing rule will not be linear. Thus, the simple commission schemes often used in practice are not the most efficient way to motivate a risk averse agent when sequential effort decisions are involved.

The presence of the additional decision information, x_1, for the agent, which is the only difference between the sequential effort problem and the effort allocation problem, introduces more complexity into the sharing rule in two ways: (i) the multiplier $\mu_2(\cdot)$ depends on x_1, and (ii) because $a_2(\cdot)$ depends on x_1, the distribution of x_2 given $a_2(\cdot)$ depends on x_1, even if $x_1|a_1$ and $x_2|a_2$ are statistically independent. The combination of these two features precludes an additively separable sharing rule. Note that $a_2(\cdot)$ depends on x_1 even if $x_1|a_1$ and $x_2|a_2$ are statistically independent. Hence, a_2's dependence on x_1 is not due to information that x_1 provides about the likelihood of x_2. Rather, the dependence is due to a wealth effect (x_1 influences the agent's position on his or her utility curve before the second effort is chosen) which the principal can use to efficiently motivate the agent. Recall that in the first best case, there is no motivational problem, and therefore the optimal a_2 does not depend on x_1 if the agent is risk averse, the principal is risk neutral, and $x_1|a_1$ and $x_2|a_2$ are statistically independent.

If, on the other hand, $x_1|a_1$ and $x_2|a_2$ are dependent, then a_2 depends on x_1 for the additional reason that x_1 provides information about the likelihood of x_2. This is true in both the first best and second best cases.

The multiplier $\mu_2(x_1)$ further complicates the sharing rule. Intuitively, it is a measure of the cost to the principal of the motivational problem for a_2. The result that $\mu_2(x_1) > 0$ for all x_1 means that no matter what the first period outcome is, the principal will not find it optimal to induce as high an effort level, a_2, as he or she could have if there were no motivational problem.

4.4 ADDITIVE EFFORT

In this section, the additive effort situation described in Section 3.5 is examined when sequential choice is allowed. The principal is assumed to be risk neutral and the agent is assumed to be risk averse. The agent is further assumed to have no intrinsic disutility for any particular task, but rather is assumed to have disutility only for the total effort expended. The agent's disutility is thus represented as $V(a_1+a_2(\cdot))$.

In first best situations, if $x_1|a_1$ and $x_2|a_2$ are independent, the principal is risk neutral, and the agent is risk averse, then the optimal $a_2(\cdot)$ does not depend on x_1 in the sequential effort case. Therefore, the first best results for the allocation of effort problem still hold for the sequential effort problem. In particular, if the means are linear in effort, that is, the means are given by ka_i, then only the sum of the efforts is of importance to the principal and the agent. If the means are given by $k_i a_i$, where $k_i \neq k_j$, $i \neq j$, then all the effort should be put into the task with the largest return per unit of effort. For more general unequal mean functions, the optimal solution will involve nonzero efforts devoted to all tasks. If the mean functions are identical nonlinear strictly increasing functions, then the optimal efforts are equal.

The second best case is quite different because of the dependence of $a_2(\cdot)$ on x_1. Recall that in the allocation problem, assuming an interior solution, the constraints require that

$$\frac{\partial EU(s(\underline{x}))}{\partial a_1} = \frac{\partial EU(s(\underline{x}))}{\partial a_2},$$

because each of the marginal expected utilities must equal the marginal disutility from the total effort, $V'(a_1+a_2)$. In the sequential effort case, the constraints become

$$\frac{\partial EU(s(\underline{x}))}{\partial a_1} = \frac{\partial EV(a_1+a_2(\cdot))}{\partial a_1} \tag{4.4.1}$$

and

$$\frac{\partial E_2 U(s(\underline{x}))}{\partial a_2} = V'(a_1+a_2(\cdot)) \quad \text{for almost every } x_1, \tag{4.4.2}$$

where $E_2U(s(\underline{x})) = \int U(s(\cdot))g(x_2|a_2(x_1))dx_2$. Equation (4.4.1) requires averaging over all possible values of x_1 and x_2, because a_1 is chosen before either outcome is available. Equation (4.4.2), on the other hand, requires averaging only over all possible values of x_2, because x_2 is the only remaining uncertainty at the time the second effort level is selected.

Corollary 4.4.1 below applies Proposition 4.2.1, which characterizes the behavior of the second stage effort strategy, to the additive effort case. Proposition 4.2.1 assumed that efforts were defined such that they were the means of the outcome distributions. In this section, efforts are assumed to be additive; assuming that efforts are simultaneously the means of the outcome distributions is overly restrictive. Therefore, Corollary 4.4.1 allows for a more general situation in which the means of the outcome distributions are functions of the efforts. This accounts for conditions on the second stage mean, $M_2(\cdot)$, in order to characterize the behavior of the second stage effort strategy. It should be noted that the definition of effort in turn influences the description of disutility captured in the disutility function $V(\cdot)$. Thus, conditions on both $V(\cdot)$ and $M_2(\cdot)$ are either implicitly or explicitly required in order to characterize the behavior of the second stage effort strategy.

<u>Corollary 4.4.1</u>. Assume that the conditions in Proposition 4.2.1 hold, except that $E(x_1|a_1) = M_1(a_1) > 0$, $E(x_2|a_2) = M_2(a_2) > 0$, and $V(a_1,a_2) = V(a_1+a_2)$, with $M_1' > 0$ and $M_2' > 0$. Let $e_1 = M_1(a_1)$ and $e_2 = M_2(a_2)$. The induced disutility function is then $V^*(e_1,e_2) = V(M_1^{-1}(e_1) + M_2^{-1}(e_2))$. If $V' > 0$, $V'' \geq 0$, $V''' \geq 0$, and $M_2'' \leq 0$, then a sufficient condition for $e_2^*(x_1)$ to be decreasing in x_1 is that $3M_2''^2 - M_2'''M_2'$ be nonnegative at a_2^*.

For example, suppose $M_1(a_1) = a_1^{\alpha}$, $M_2(a_2) = a_2^{\beta}$, and $V(a_1+a_2) = (a_1+a_2)^2$, where $0 < \alpha < 1$, $0 < \beta < 1/2$, and $a_i \geq 0$ for i=1,2. Then

$a_1 = M_1^{-1}(e_1) = e_1^{1/\alpha}$ and $a_2 = M_2^{-1}(e_2) = e_2^{1/\beta}$. Corollary 4.4.1 shows that $e_2^*(x_1)$ is decreasing in x_1, since $V' = 2(a_1+a_2) > 0$ for $a_1+a_2 \neq 0$, $V'' = 2 > 0$, $V''' = 0$, $M_2'' = \beta(\beta-1)a_2^{\beta-2} < 0$, and $3M_2''^2 - M_2'''M_2' = 3\beta^2(\beta-1)^2 a_2^{2\beta-4} - \beta^2(\beta-1)(\beta-2)a_2^{2\beta-4} > 0$ for $\beta < \frac{1}{2}$. By Proposition 4.2.1, the agent's expected second stage net utility is therefore increasing in x_1.

The remainder of this section makes Pareto comparisons between effort strategies in the sequential effort model, in which information becomes available at a fixed point during the period. In the discussion that follows, the agent has effort choices for two tasks before any information is observed. After observing the information (if a nonzero effort is exerted at a task, then the associated outcome is observed), the agent can choose to begin, continue, or discontinue exerting effort at the two tasks. Letting the first subscript on a and on x denote time (the stage) and letting the second subscript denote the tasks, the sequential choice scenario under discussion can be depicted as follows:

a_{11} and/or a_{12} exerted	Information Point: x_{11} and/or x_{12} observed	$a_{21}(\cdot)$ and/or $a_{22}(\cdot)$ exerted	x_{21} and/or x_{22} observed

In the allocation of effort situation discussed in Chapter 3, the agent's effort decisions are made and the efforts are exerted before the outcomes are observed. The situation may be viewed as one in which the efforts are exerted sequentially, or simultaneously. In either case, the outcomes are observed only after both efforts have been exerted. The allocation of effort case can be thought of as a special case of the situation described above, where the null information system is in effect at the Information Point. Neither x_{11} nor x_{12} is observed until the end of the period. The efforts a_{21} and a_{22} are thus independent of x_{11} and x_{12}. At the end of the period, $a_1 = a_{11} + a_{21}$ and $a_2 = a_{12} + a_{22}$ will have been exerted, and $x_1 = x_{11} + x_{21}$ and $x_2 = x_{12} + x_{22}$ will have been observed. Analysis similar to that in the proofs of the propositions in Section 3.5 establishes the sequential effort results below. Part (i) of Proposition 4.4.2 deals with situations with effort devoted to only one task at a time, while parts (ii) and (iii) deal with situations in which effort is devoted to more than one task at a time.

Proposition 4.4.2. Suppose the principal is risk neutral. Suppose further that the agent's utility for wealth is the square root utility function and that the agent's disutility is a function of the total effort expended. Finally, suppose that for i,j = 1,2, the x_{ij}'s given the corresponding a_{ij}'s are independent, and $E(x_{ij}|a_{ij}) = ka_{ij}$, where k is a constant.

(i) If it is optimal for the principal to induce (1) $a_{11} > 0$, $a_{12} = 0$, $a_{21}(x_{11}) > 0$, and $a_{22}(x_{11}) \equiv 0$, then it is also optimal for the principal to induce (2) $a_{11} > 0$, $a_{12} = 0$, $a_{21}(x_{11}) \equiv 0$, and $a_{22}(x_{11}) > 0$, or to induce (3) $a_{11} = 0$, $a_{12} > 0$, $a_{21}(x_{12}) > 0$, and $a_{22}(x_{12}) \equiv 0$, or to induce (4) $a_{11} = 0$, $a_{12} > 0$, $a_{21}(x_{12}) \equiv 0$, and $a_{22}(x_{12}) > 0$. That is, if one of the four combinations of efforts (1) through (4) is optimal, then the principal is indifferent among the four combinations. This result holds no matter what the risk averse agent's utility for wealth is. Moreover, means that are linear in effort are not required.

(ii) If $x_{ij}|a_{ij}$ is normally distributed with mean ka_{ij} and variance σ^2, then

(a) the best effort strategy with $a_{11} > 0$, $a_{12} = 0$, $a_{21}(x_{11}) > 0$, and $a_{22}(x_{11}) > 0$ is Pareto inferior to some effort strategy with $a_{11} > 0$, $a_{12} = 0$, $a_{21}(x_{11}) > 0$, and $a_{22}(x_{11}) \equiv 0$, and

(b) the best effort strategy with $a_{11} > 0$, $a_{12} > 0$, $a_{21}(x_{11},x_{12}) > 0$, and $a_{22}(x_{11},x_{12}) > 0$ is Pareto inferior to some effort strategy with $a_{11} > 0$, $a_{12} > 0$, $a_{21}(x_{11},x_{12}) \equiv 0$, and $a_{22}(x_{11},x_{12}) > 0$.

(iii) If $x_{ij}|a_{ij}$ is exponentially distributed with mean ka_{ij}, then

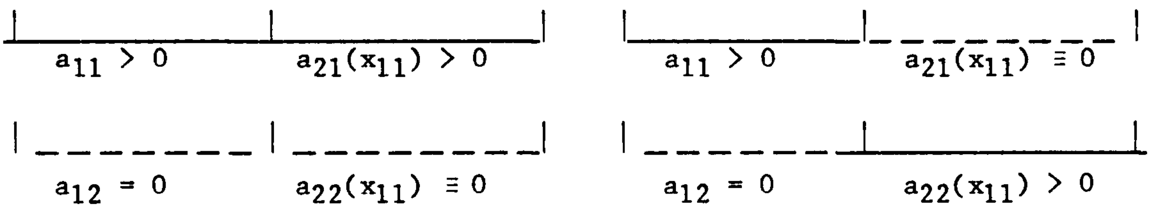
(a) the best effort strategy with $a_{11} > 0$, $a_{12} = 0$, $a_{21}(x_{11}) > 0$, and $a_{22}(x_{11}) \equiv 0$ is Pareto inferior to some effort strategy with $a_{11} > 0$, $a_{12} = 0$, $a_{21}(x_{11}) > 0$, and $a_{22}(x_{11}) > 0$, and

(b) the best effort strategy with $a_{11} > 0$, $a_{12} > 0$, $a_{21}(x_{11},x_{12}) \equiv 0$, and $a_{22}(x_{11},x_{12}) > 0$ is Pareto inferior to some effort strategy with $a_{11} > 0$, $a_{12} > 0$, $a_{21}(x_{11},x_{12}) > 0$, and $a_{22}(x_{11},x_{12}) > 0$.

The results in Proposition 4.4.2 can be depicted as follows, where solid lines indicate nonzero effort, and dashed lines indicate zero (no) effort. The first line in each pair of lines represents the first task, and the second line in each pair represents the second task.

(i)

(1) (2)



The principal is indifferent between (1) and (2). Alternative (4) is similar to alternative (2), with the tasks renumbered, and alternative (3) is similar to alternative (1).

(ii) (a)

(A)

$a_{11} > 0$ $\quad$ $a_{21}(x_{11}) > 0$

$a_{12} = 0$ $\quad$ $a_{22}(x_{11}) > 0$

(B)

$a_{11} > 0$ $\quad$ $a_{21}(x_{11}) > 0$

$a_{12} = 0$ $\quad$ $a_{22}(x_{11}) > 0$

The principal prefers some form of (B) to the best possible form of (A).

(ii) (b)

(C)

$a_{11} > 0$ $\quad$ $a_{21}(x_{11},x_{12}) > 0$

$a_{12} > 0$ $\quad$ $a_{22}(x_{11},x_{12}) > 0$

(D)

$a_{11} > 0$ $\quad$ $a_{21}(x_{11},x_{12}) \equiv 0$

$a_{12} > 0$ $\quad$ $a_{22}(x_{11},x_{12}) > 0$

The principal prefers some form of (D) to the best possible form of (C). The results in (ii) say that whether effort is exerted at one or two tasks initially, all effort should be concentrated in only one task at the second stage. Because of the assumed independence, it does not matter which task is chosen.

In part (iii) of the proposition, the results in (ii)(a) and (b) are reversed. That is, whether effort is exerted at one or two tasks initially, effort should be split across two tasks at the second stage. It is preferable to induce the agent to diversify effort after receipt of the information x_1 when the outcomes are exponentially distributed as described, and it is preferable not to induce the agent to diversify effort when the outcomes are normally distributed as described. In part (i) of the proposition, diversification of effort is not in question. Because the outcomes conditional on the efforts are independent and identically distributed, the principal is indifferent among the four alternatives (1) through (4).

As in Section 3.5, the results in parts (ii) and (iii) are partly explainable in terms of the variances of the total outcomes. For simplicity, consider a comparison between a fixed amount of effort, a, devoted to only one task, or divided across two tasks. Let x_1 and x_2 denote the outcomes of the two tasks, and let ka_1 and ka_2 denote their respective means, where a_i is the effort devoted to task i. Since the means of each of the individual outcomes are linear in effort, the total effort expended is the only quantity of relevance for the purpose of comparing the means of the total outcomes (x_1 if effort is devoted only to one task, and x_1+x_2 if effort is devoted to two tasks). For the normal distribution in part (ii) of Proposition 4.4.2, $\mathrm{Var}(x_1|a_1=a) = \sigma^2$, and $\mathrm{Var}(x_1+x_2|a_1+a_2=a) = 2\sigma^2$. For the exponential distributions in part (iii), $\mathrm{Var}(x_1|a_1=a) = k^2a^2$, and $\mathrm{Var}(x_1+x_2|a_1+a_2=a) < k^2a^2$. For the normal distribution, the variance of the total outcome is smaller when all the effort is devoted to one task, while for the exponential distribution, the variance of the total outcome is smaller when all the effort is divided across two tasks. This observation can be related to the information content of the outcomes considered as signals about the agent's effort(s). The quantity $I(a) = \int f^2(x|a)/f(x|a)dx$, called Fisher's information about the parameter a contained in the data (see, for example, Cox and Hinkley, 1974), is used as a measure of information content about a in x. For both the normal and exponential distributions described above, I(a) is the reciprocal of the variance. Thus, for the normal case, there is "more" information about the agent's effort when all effort is devoted to one task than there is when the effort is divided across the tasks. The opposite is true for the exponential distribution.

Proposition 4.4.2 does not state what the optimal effort strategies are in each case. The comparisons in parts (ii) and (iii) are between situations with the same information available at the beginning of the second stage. For example, in (ii)(a), the comparison is between two situations in which only x_{11} is available at the beginning of the second stage. Comparisons of situations with differing information available at the beginning of the second stage are more difficult to make.

4.5 SUMMARY AND DISCUSSION

This chapter examined the problem of sequential effort decisions within one period. The sequential aspect arose because the agent observed an outcome affected by the first effort choice before making the second effort choice, which affected a second outcome. The agent was paid only after both efforts were exerted and both outcomes were observed.

In the first best case, the characterization of the optimal sharing rule in the sequential effort case is similar in spirit to that in the allocation of effort case. That is, if one person is risk neutral and the other is risk averse, then the risk neutral person bears the risk. If both the principal and the agent are risk averse, then the risk is shared, with the sharing rule a function of the sum of the outcomes.

The first best characterization of the optimal efforts is different in the sequential effort case than in the allocation of effort case. The second effort choice may now depend on the first outcome and the first effort choice. If both of the individuals are risk averse, then the optimal second stage effort strategy will depend on x_1, the first outcome. The second stage effort strategy will also depend on x_1 if the joint density of the two outcomes given the actions is $f(x_1|a_1)g(x_2|x_1,a_1,a_2(\cdot))$. However, if at least one of the individuals is risk neutral and the joint density of the two outcomes given the actions is $f(x_1|a_1)g(x_2|x_1,a_1,a_2(\cdot))$, then the optimal second stage effort strategy will be independent of x_1.

The second stage effort strategy may depend on x_1 because of a "wealth" ("risk aversion") effect, or because of an "information" effect. The wealth effect occurs when both individuals are risk averse, because a risk averse individual's marginal utility varies at different points of the utility curve. The first outcome determines where on the utility curve the individual is, so the individual will want the second stage effort adjusted according to the value of the first outcome. The information effect occurs when the two outcomes are dependent. Depending on the nature of the correlation between the two outcomes, the principal may wish to induce the agent to choose the second stage effort strategy to be an increasing or decreasing function of the first outcome. Proposition 4.1.1 provides a precise expression for the derivative of the second stage effort strategy with respect to the first outcome.

The analysis in the second best case allowed for nonindependence of the outcomes. As usual, the principal was assumed to be risk neutral and the agent was assumed to be risk averse. The characterization of the optimal sharing rule in the sequential effort case is similar to the characterization in the allocation of effort case, except that the multipler μ_2 and the effort strategy a_2 may depend on x_1. Although in general, a_2 will depend on x_1, a_2 will be independent of x_1 if the agent is risk neutral and the joint density of the outcomes is of the form $f(x_1|a_1)g(x_2|a_2(\cdot))$.

Proposition 4.2.1 assumed a square root utility function for the agent and conditionally independent outcomes given the actions. It was shown that the agent's second stage effort strategy will be decreasing in x_1. Intuitively, this is because the higher x_1 is, the more costly it is for the principal to induce any particular level of a_2. The agent's decreasing marginal utility for wealth and increasing marginal disutility for effort account for the increasing costliness of inducing a_2. Since these characteristics hold in general, the results in Proposition 4.2.1 should hold for other utility functions.

The case of perfectly correlated outcomes was next examined. A sharing rule which incorporates a penalty wage for the second stage was shown to generally result in a second effort strategy that is decreasing in the first outcome if the state is random noise. If the state reveals information about the production technology, then the second effort strategy may be strictly increasing in the first outcome. The information effect of the first outcome in the perfectly correlated case can therefore override the wealth effect, changing the behavior of the agent's effort strategy. The behavior of a_2 is more complex when the outcomes are imperfectly correlated.

It was next shown that the conditions which guaranteed an optimal sharing rule that is additively separable in the outcomes in the allocation of effort case will not guarantee an additively separable sharing rule in the sequential effort case. Thus, the presence of the additional decision information, x_1, precludes an additively separable sharing rule.

The first best additive effort results for the allocation of effort problem analyzed in Section 3 also hold for the sequential effort case. The second best results in the sequential case differ from those in the allocation case because of the role that the first outcome plays as pre-decision information for the second effort choice. Corollary 4.4.1 provides conditions under which the agent's second stage effort strategy will be decreasing in x_1, the first outcome, when effort is additive. Finally, the boundary versus interior solution results in Section 3.5 were applied to the sequential effort case in order to obtain Pareto comparisons between effort strategies with varying degrees of diversification of effort. The results were related to Fisher's information statistic, a measure of the information content of the outcome about the agent's effort.

The conditional investigation problem is somewhat related to the sequential effort problem. In the conditional investigation problem, the agent exerts effort, and both the principal and the agent observe the outcome x. The principal then has the option of observing y, an additional signal about the agent's effort. The agent's compensation is s(x) or t(x,y), depending on what was jointly observed. Cost variance investigation, a familiar problem in accounting, has been modeled as a conditional investigation problem (see, for example, Baiman and Demski, 1980a,b) in which x is a cost and y is the result of an investigation to try to determine the reason for the cost's deviation from a preset standard. The problem is similar to the sequential effort problem in that decisions are based on an initial outcome. However, after the initial outcome, the principal chooses an act in the conditional investigation problem, and the agent chooses an act in the sequential effort problem. The major focus in the conditional investigation problem has been on the determination of the optimal investigation strategy; such a question is not at all relevant in the sequential effort choice problem. Some additional comments about the conditional investigation problem will be made in the next chapter.

As remarked at the end of Chapter 3, the sequential effort case can be viewed as a special case of the two-period agency problem in which the principal's and the agent's expected utilities depend only on the total return over the entire time horizon. Thus, the sequential effort results have potential applications in such multiperiod situations.

CHAPTER 5

SUGGESTED FURTHER RESEARCH

This chapter concludes the thesis with suggestions for further research. The first section discusses possible extensions to theoretical agency results, and the second section discusses possible applications of the agency theory results to a traditional accounting topic, cost variance investigation.

5.1 THEORETICAL AGENCY EXTENSIONS

A number of generalizations of the results in this thesis are desirable. For example, in the allocation of effort setting with additive effort, it is desirable to obtain results for a more general class of utility functions and for nonindependent distributions of incomes. A similar remark holds for some of the results in the sequential effort setting. The situation with multiple agents was discussed briefly in Section 3.6, where the agents were salespeople in a firm. The important problem of collusion among agents in order to conceal shirking or the theft of assets has largely been unexplored. Beck (1982), however, has recently taken an incentive contracting approach to the problem of collusion for the purpose of concealing the theft of assets.

As remarked earlier, many accounting and other business issues are best addressed in a multiperiod setting. Lambert (1981, 1983) has analyzed a special case of the multiperiod agency problem in which utilities are additive over time and the outcomes are independent. Chapter 4 of this thesis analyzed a different special case of the multiperiod problem. The analysis allows for nonindependent outcomes, and assumes that the agent is paid only at the end of the time horizon, even though the effort choices and the observations of the outcomes are sequential. The analysis is thus suitable for short-term horizons in which the principal and the agent are concerned only with their total shares at the end of the time horizon. Results for more general multiperiod situations are desirable. These situations are, of course, more difficult to analyze.

5.2 APPLICATION TO VARIANCE INVESTIGATION

A great deal of attention has been focused on strategies for investigating the underlying causes of cost variances or deviations from standards. Most of the analytical research has assumed that investigations reveal the state of a mechanistic production process, and that the investigator can return an "out-of-control" state to an "in-control" state (Kaplan, 1975). Thus, only the correctional purposes of investigations were examined. Correctional benefits occur, for example, when costs are higher for a malfunctioning machine than for a properly functioning machine.

In some situations, the primary focus is on evaluating a manager who has control over a mechanistic process. In such situations, there may be motivational as well as correctional benefits to investigating variances. The manager's actions can be influenced by the possibility of an investigation if a reward or penalty is based on the results of the investigation. The motivational purposes of investigations have recently come to attention in the analytical literature. Baiman and Demski (1980a, 1980b) have explored the motivational aspects of variance analysis procedures in a one-period agency model, with a single-dimensional effort variable. In both of the analyses, the agent is responsible for a production process which generates a monetary outcome determined by the agent's effort and some exogenous randomness. The monetary outcome, owned by the principal, is assumed to be jointly observable, while the agent's effort is not. The principal can, however, conduct a costly investigation in order to obtain a further imperfect signal which is independent of the outcome but informative about the agent's effort. The nature of the investigation strategy was characterized, and the use of the information for motivational purposes was demonstrated. Lambert (1984) extended the analysis by allowing for a nonindependent additional signal about the agent's effort, and showed that the investigation strategy would differ from that obtained by Baiman and Demski.

A number of extensions to the Baiman-Demski analysis are possible. One extension is to allow for multiple effort decisions by the agent. Feltham and Matsumura (1979), for example, suggested three different effort decisions the agent might be responsible for: 1) bringing the system back into control at the beginning of the period after detecting that it is out of control; 2) keeping the process in control during the period given that the process is in control at the beginning of the period; 3) influencing or controlling the operating costs or the outcome during the period. Their analysis did not focus explicitly on the tradeoffs between the efforts expended by the agent. Instead, the focus was on characterizing the optimal investigation strategy and sharing rule for an infinite-horizon Markov process.

Another extension to the Baiman-Demski analysis is the extension to multiple periods. One approach would be to extend the analysis to a finite-horizon model. Another approach would be to extend the analysis to an infinite-horizon model. It has been argued that in finite-horizon multiperiod problems involving two players, the factor that overshadows all others is the players' knowledge that they have arrived at the last play. When the players expect that there will always be another "play" of the game, the appropriate concept is the repeated game, in which there are an infinite number of plays of the single game (Rubinstein, 1979).

BIBLIOGRAPHY

Amershi, A., "Valuation of Performance Information Systems in Agencies," Faculty of Commerce and Business Administration, University of British Columbia, working paper (1982).

_____ and J. Butterworth, "Pareto and Core Syndicates: Risk-Sharing in Perspective," Faculty of Commerce and Business Administration, University of British Columbia, working paper (1981).

Baiman, S., "Agency Research in Managerial Accounting: A Survey," Journal of Accounting Literature (Spring 1982), 154-213.

_____ and J. Demski, "Variance Analysis Procedures as Motivational Devices," Management Science (August 1980), 840-848.

_____ and _____, "Economically Optimal Performance Evaluation and Control Systems," Journal of Accounting Research Supplement (1980), 184-220.

_____ and Evans, John H. III, "Pre-Decision Information and Participative Management Control Systems," Journal of Accounting Research (Autumn 1983), 371-395.

Beck, P., "Internal Control and Irregularities: An Incentive Contracting Model," Graduate School of Management, University of California, Los Angeles, working paper (November 1982).

Berger, P., "Optimal Compensation Plans: The Effect of Uncertainty and Attitude Toward Risk on the Salesman Effort Allocation Decision," in E. Mazze (ed.), Proceedings of the 1975 Marketing Educator's Conference. American Marketing Association: Chicago (1975), 517-520.

Blackwell, D., "Equivalent Comparison of Experiments," Annals of Mathematical Statistics (1953), 265-272.

Christensen, John, "Communication in Agencies," Bell Journal of Economics (Autumn 1981), 661-674.

_____, "The Determination of Performance Standards and Participation," Journal of Accounting Research (Autumn 1982, Part II), 589-603.

Cox, D. and D. Hinkley, Theoretical Statistics. Chapman and Hall: London (1979).

DeGroot, M., Optimal Statistical Decisions. McGraw-Hill: New York (1970).

Demski, J. and G. Feltham, Cost Determination: A Conceptual Approach. Iowa State University Press: Ames (1976).

_____ and _____, "Economic Incentives in Budgetary Control Systems," The Accounting Review (April 1978), 336-359.

_____ and D. Sappington, "Delegated Expertise," Graduate School of Business, Stanford University, working paper (1983).

Farley, J., "An Optimal Plan for Salesmen's Compensation," Journal of Marketing Research (May 1964), 39-43.

Fellingham, J., Y. Kwon, and D. Newman, "Ex Ante Randomization in Agency Models," Graduate School of Business, University of Texas at Austin, working paper #82/83-1-12 (February 1983).

Feltham, G., "Cost Aggregation: An Information Economic Analysis," Journal of Accounting Research (Spring 1977), 42-70.

_____, "Optimal Incentive Contracts: Penalties, Costly Information and Multiple Workers," Faculty of Commerce and Business Administration, University of British Columbia, working paper #508 (October 1977).

_____ and E. Matsumura, "Cost Variance Investigation: An Agency Theory Perspective," Faculty of Commerce and Business Administration, University of British Columbia, working paper (July 1979).

Gjesdal, F., "Stewardship Accounting: Controlling Informational Externalities," Ph.D. dissertation, Stanford University (1978).

_____, "Accounting for Stewardship," Journal of Accounting Research (Spring 1981), 208-231.

_____, "Information and Incentives: The Agency Information Problem," Review of Economic Studies (1982), 373-390.

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

Holmstrom, B., "On Incentives and Control in Organizations," Ph.D. dissertation, Stanford University (1977).

_____, "Moral Hazard and Observability," Bell Journal of Economics (Spring 1979), 74-91.

_____, "Moral Hazard in Teams," Bell Journal of Economics (Spring 1982), 324-340.

Horngren, C., Cost Accounting: A Managerial Emphasis (Fifth edition). Prentice Hall: Englewood Cliffs (1982).

Itami, H., "Analysis of the Optimal Linear Goal-Based Incentive System," Department of Commerce, Hitotsubashi University, working paper (March 1979).

Johnson, N. and S. Kotz, Continuous Multivariate Distributions. Wiley: New York (1972).

Kaplan, R., "The Significance and Investigation of Cost Variances: Survey and Extensions," Journal of Accounting Research (Autumn 1975), 311-337.

Lal, R., "A Theory of Salesforce Compensation Plans," Ph.D. dissertation, GSIA, Carnegie-Mellon University (1982).

Lambert, R., "Managerial Incentives in Multiperiod Agency Relationships," Ph.D. dissertation, Stanford University (1981).

_____, "Long-Term Contracts and Moral Hazard," Bell Journal of Economics (Autumn 1983), 441-452.

_____, "Variance Investigation in Agency Settings," Kellogg Graduate School of Management, Northwestern University, working paper (1984).

Lehmann, E., Testing Statistical Hypotheses. John Wiley & Sons, Inc.: New York (1959).

Mirrlees, J., "Notes on Welfare Economics, Information and Uncertainty," in M. Balch, D. McFadden, and S. Wu (eds.), Essays on Economic Behavior Under Uncertainty. North-Holland (1974).

Peng, J., "Simultaneous Estimation of the Parameters of Independent Poisson Distribution," Ph.D. dissertation, Technical Report #78, Department of Statistics, Stanford University (1975).

Radner, R. and M. Rothschild, "On the Allocation of Effort," Journal of Economic Theory (1975), 358-376.

Rubinstein, A., "Offenses That May Have Been Committed by Accident - an Optimal Policy of Retribution," in S. Brams, A. Schotter, and G. Schrodiauer (eds.), Applied Game Theory. Physica-Verlag: Wurzburg (1979).

Shavell, S., "Risk Sharing and Incentives in the Principal and Agent Relationship," Bell Journal of Economics (Spring 1979), 55-73.

Steinbrink, J., "How to Pay Your Sales Force," Harvard Business Review (July-August 1978), 111-122.

Stiglitz, J., "Incentives, Risk, and Information: Notes Toward a Theory of Hierarchy," Bell Journal of Economics (Autumn 1975), 552-579.

Weinberg, C., "An Optimal Commission Plan for Salesmen's Control Over Price," Management Science (April 1975), 937-943.

_____, "Jointly Optimal Sales Commissions for Nonincome Maximizing Salesforces," Management Science (August 1978), 1252-1258.

Zimmerman, J., "The Costs and Benefits of Cost Allocations," The Accounting Review (July 1979), 504-521.

Appendix 1

Table II
One-parameter Exponential Family Q

$f(x|a) = \exp[z(a)x - B(z(a))]h(x)$

	Exponential	Normal	Gamma	Poisson	Binomial
$f(x\|a)$	$\frac{1}{M(a)} \exp[\frac{-x}{M(a)}]$	$\frac{1}{\sqrt{2\pi}} \exp[\frac{-(x-M(a))^2}{2\sigma^2}]$	$\frac{\lambda^n x^{n-1} e^{-\lambda x}}{\Gamma(n)}$	$\exp[-M(a)]\frac{(M(a))^x}{x!}$	$\binom{n}{x}(\frac{M(a)}{n})^x (1 - \frac{M(a)}{n})^{n-x}$
$E(x\|a)$	$M(a)$	$M(a)$	$M(a)$ $(= n/\lambda)$	$M(a)$	$M(a)$
$Var(x\|a)$	$M^2(a)$	σ^2	$\frac{M^2(a)}{n}$	$M(a)$	$M(a)\ [1 - \frac{M(a)}{n}]$
$z(a)$	$-\frac{1}{M(a)}$	$\frac{M(a)}{\sigma^2}$	$-\frac{n}{M(a)}$	$\ln M(a)$	$\ln M(a) - \ln (n-M(a))$
$B(z)$	$-\ln (-z)$	$\frac{\sigma^2}{2} z^2$	$n \ln (-\frac{n}{z})$	$\exp (z)$	$nz - n \ln[\frac{ne^z}{1 + e^z}]$
$B'(z)$	$-\frac{1}{z}$	$\sigma^2 z$	$-\frac{n}{z}$	$\exp (z)$	$\frac{ne^z}{1 + e^z}$
$B''(z)$	$\frac{1}{z^2}$	σ^2	$\frac{n}{z^2}$	$\exp (z)$	$\frac{ne^z}{(1 + e^z)^2}$
f_a/f	$\frac{M'(a)}{M^2(a)}(x-M(a))$	$\frac{M'(a)}{\sigma^2}(x-M(a))$	$\frac{nM'(a)}{M^2(a)}(x-M(a))$	$\frac{M'(a)}{M(a)}(x-M(a))$	$[\frac{M'(a)}{M(a)} + \frac{M'(a)}{n-M(a)}]\ (x-M(a))$

Note: The exponential distribution is a special case of the gamma distribution but is listed separately because of its wide use.

The following calculations for the one-parameter exponential family, Q, given in Table II, will be useful in the proofs of the results in Chapters 3 and 4.

$$\frac{f_a}{f} = \frac{d \ln f}{da} = \frac{d}{da}\,[z(a)x - B(z(a))]$$

$$= z'(a)x - B'(z(a))z'(a)$$

$$= z'(a)[x - B'(z(a))]$$

$$= z'(a)(x - E(x|a)). \tag{A1.1}$$

$$\int \frac{f_a^2(x|a)}{f(x|a)}dx = \int z'(a)(x-E(x|a))f_a(x|a)dx$$

$$= z'(a)\frac{d}{da}\int xf(x|a)dx - z'(a)E(x|a)\int f_a(x|a)dx$$

$$= z'(a)\frac{d}{da}\,B'(z(a)), \quad \text{since } E(x|a) = B'(z(a))$$

$$\text{and } \int f_a(x|a)dx = 0$$

$$= (z'(a))^2 B''(z(a)). \tag{A1.2}$$

$$\int \frac{f_a^3(x|a)}{f^2(x|a)}\,dx\,\Bigg|_{a^*} \equiv \frac{d}{da}\int \left(\frac{f_a}{f}\Bigg|_{a^*}\right)^2 f(x|a)dx\,\Bigg|_{a^*}$$

$$= \frac{d}{da}(z'(a^*))^2 \int (x-E(x|a^*))^2 f(x|a)dx\,\Bigg|_{a^*}.$$

Note that $\int (x-E(x|a^*))^2 f(x|a)dx$

$$= \int (x-E(x|a) + E(x|a) - E(x|a^*))^2 f(x|a)dx$$

$$= \int (x-E(x|a))^2 f(x|a)dx + (E(x|a) - E(x|a^*))^2$$

$$- 2(E(x|a) - E(x|a^*)) \int (x-E(x|a))f(x|a)dx$$

$$= Var(x|a) + (E(x|a) - E(x|a^*))^2.$$

Therefore,

$$\int \frac{f_a^3(x|a)}{f^2(x|a)}\Bigg|_{a*} dx = (z'(a*))^2 \frac{d}{da}(B''(z(a)))\Bigg|_{a*}$$

$$= (z'(a*))^3 B'''(z(a*)). \tag{A1.3}$$

$$\int \frac{f_a(x|a)}{f(x|a)}\Bigg|_{a*} f_{aa}(x|a)dx\Bigg|_{a*} \equiv z'(a*)\frac{d^2}{da^2}\int (x-E(x|a*))f(x|a)dx\Bigg|_{a*}$$

$$= z'(a*)\frac{d^2}{da^2}(E(x|a) - E(x|a*))\Bigg|_{a*}$$

$$= z'(a*)\frac{d^2}{da^2}B'(z(a))\Bigg|_{a*}$$

$$= z'(a*)\frac{d}{da}(B''(z(a))z'(a))\Bigg|_{a*}$$

$$= z'(a*)[B'''(z(a*))(z'(a*))^2 + B''(z(a*))z''(a*)]. \tag{A1.4}$$

Second Best, Additive Effort Example (Section 3.5)

Suppose the principal is risk neutral,

$$\phi(x_1,x_2|a_1,a_2) = f(x_1|a_1)f(x_2|a_2) \;,$$

$$V(a_1,a_2) = V(a_1 + a_2) \;, \text{ and}$$

$$U^A(s) = 2\sqrt{s} \;.$$

Then $s*(x_1,x_2) = R^2(\underline{a}*)$, where

$$R(\underline{a}*) = \lambda + \sum_{i=1}^{2} \mu_i \frac{f_{a_i}(x_i|a_i)}{f(x_i|a_i)}\Bigg|_{\underline{a}*} \;.$$

Let $K(s,\underline{a}) = \int 2\, R(\underline{a}^*)\,\phi(\underline{x}|\underline{a})\, d\underline{x}$. The agent's expected utility is

$$EU(s^*,\underline{a}) = K(s^*,\underline{a}) - V(a_1 + a_2)$$

and the principal's expected utility is

$$G(s^*,\underline{a}) = \int (x_1 + x_2 - R^2(\underline{a}^*))\;\phi(\underline{x}|\underline{a})\, dx.$$

The Hamiltonian, with $s^*(\underline{x})$, is

$$H = G(s^*,\underline{a}) + \lambda(EU(s^*,\underline{a}) - \bar{u}) + \sum_{k=1}^{2} \mu_k \frac{\partial EU(s^*,\underline{a})}{\partial a_k} .$$

The first order conditions are

1) $$\left.\frac{\partial G}{\partial a_j}\right|_{\underline{a}^*} + \lambda\cdot 0 + \left.\sum_{i=1}^{2} \mu_i \frac{\partial^2 EU}{\partial a_i \partial a_j}\right|_{\underline{a}^*} = 0, \; j=1,2.$$

That is, $G_{a_1} + \mu_1(K_{a_1a_1} - V'') + \mu_2(K_{a_1a_2} - V'') = 0$

and $G_{a_2} + \mu_1(K_{a_1a_2} - V'') + \mu_2(K_{a_2a_2} - V'') = 0$,

which imply that

$$G_{a_1} + \mu_1 K_{a_1a_1} + \mu_2 K_{a_1a_2} = G_{a_2} + \mu_1 K_{a_1a_2} + \mu_2 K_{a_2a_2} , \tag{A1.5}$$

where all the functions are evaluated at $\underline{a}^* = (a_1^*,a_2^*)$.

2) $\frac{\partial EU}{\partial a_j} = 0$, $j=1,2$. I.e.,

$$\frac{\partial}{\partial a_1} \int 2R(\underline{a}^*)\; f(x_1|a_1)\; f(x_2|a_2)\; dx_1 dx_2 - V'(a_1 + a_2) = 0$$

and $$\frac{\partial}{\partial a_2} \int 2R(\underline{a}^*)\; f(x_1|a_1)\; f(x_2|a_2)\; dx_1 dx_2 - V'(a_1 + a_2) = 0 .$$

Thus, $$\frac{\partial}{\partial a_1} \int [\lambda + \sum_{i=1}^{2} \mu_i \left.\frac{f_{a_i}(\cdot)}{f(\cdot)}\right|_{\underline{a}^*}] f(x_1|a_1)\ f(x_2|a_2)\ dx_1 dx_2$$

$$= \frac{\partial}{\partial a_2} [\lambda + \sum_{i=1}^{2} \mu_i \left.\frac{f_{a_i}(\cdot)}{f(\cdot)}\right|_{\underline{a}^*}] f(x_1|a_1)\ f(x_2|a_2)\ dx_1 dx_2 ,$$

which implies that

$$\frac{\partial}{\partial a_1} \int \mu_1 \left.\frac{f_{a_1}}{f}\right|_{\underline{a}^*} f(x_1|a_1)\ dx_1 = \frac{\partial}{\partial a_2} \int \mu_2 \left.\frac{f_{a_2}}{f}\right|_{\underline{a}^*} f(x_2|a_2)\ dx_2$$

$$\text{since } \frac{\partial}{\partial a_j} \int \mu_i \frac{f_{a_i}}{f} f(x_i|a_i)\ dx_i = 0 \qquad \text{for } i \neq j.$$

Therefore, $$\mu_1 \int \left.\frac{f_{a_1}^2}{f}\right|_{\underline{a}^*} dx_1 = \mu_2 \int \left.\frac{f_{a_2}^2}{f}\right|_{\underline{a}^*} dx_2. \tag{A1.6}$$

Let J denote the quantity in (A1.6).

G_{a_1} can be written as

$$\frac{\partial}{\partial a_1} [E(x_1|a_1) + E(x_2|a_2)] - \frac{\partial}{\partial a_1} \int [\lambda^2 + 2\lambda\mu_1 \left.\frac{f_{a_1}}{f}\right|_{\underline{a}^*} + \mu_1^2 \{ \left.\frac{f_{a_1}}{f}\right|_{\underline{a}^*} \}^2$$

$$+ 2\lambda\mu_2 \left.\frac{f_{a_2}}{f}\right|_{\underline{a}^*} + 2\mu_1\mu_2 \left[\frac{f_{a_1}}{f} \cdot \frac{f_{a_2}}{f} \right]\Big|_{\underline{a}^*}$$

$$+ \mu_2^2 \{ \left.\frac{f_{a_2}}{f}\right|_{\underline{a}^*} \}^2] \cdot f(x_1|a_1) f(x_2|a_2) dx_1 dx_2$$

$$= \frac{\partial E(x_1|a_1)}{\partial a_1} - [\ 2\lambda\mu_1 \frac{\partial}{\partial a_1} \int \frac{f_{a_1}}{f}\Bigg|_{\underline{a}^*} f(x_1|a_1)dx_1$$

$$+ \mu_1^2 \frac{\partial}{\partial a_1} \int \{ \frac{f_{a_1}}{f}\Bigg|_{\underline{a}^*} \}^2 f(x_1|a_1)dx_1$$

$$+ 2\mu_1\mu_2 \frac{\partial}{\partial a_1} \int (\frac{f_{a_1}}{f} \cdot \frac{f_{a_2}}{f})\Bigg|_{\underline{a}^*} f(x_1|a_1)\ f(x_2|a_2)dx_1dx_2\ ].$$

The last term in the sum is 0 when evaluated at $\underline{a}^*$ because x_1 and x_2 are conditionally independent and

$$\int \frac{f_{a_2}(x_2|a_2^*)}{f(x_2|a_2^*)} \cdot f(x_2|a_2^*)dx_2 = 0.$$

Thus, at $\underline{a}^*$,

$$G_{a_1} = \frac{\partial E(x_1|a_1)}{\partial a_1}\Bigg|_{\underline{a}^*} - 2\lambda J - \mu_1^2 \int \frac{f_{a_1}^3}{f^2}\Bigg|_{\underline{a}^*} dx_1.$$

Now $K_{a_1a_1} = \frac{\partial^2}{\partial a_1^2} \{ 2[\ \lambda + \mu_1 \int \frac{f_{a_1}}{f}\Big|_{\underline{a}^*} f(x_1|a_1)dx_1 + \mu_2 \int \frac{f_{a_2}}{f}\Big|_{\underline{a}^*} f(x_2|a_2)dx_2\ ]\}$

$$= 2\mu_1 \int \frac{f_{a_1}}{f}\Bigg|_{\underline{a}^*} f_{a_1a_1}(x_1|a_1)dx_1$$

and $K_{a_1a_2} = K_{a_2a_1} = \frac{\partial}{\partial a_1} [\ 2\mu_2 \int \frac{f_{a_2}}{f}\bigg|_{\underline{a}^*} f(x_2|a_2)dx_2\ ] = 0$.

Therefore, (A1.5) can be written as

$$\frac{\partial E(x_1|a_1)}{\partial a_1}\bigg|_{\underline{a}^*} - 2\lambda J - \mu_1^2 \int \frac{f_{a_1}^3}{f^2}\bigg|_{\underline{a}^*} dx_1 + 2\mu_1^2 \int \frac{f_{a_1}}{f}\bigg|_{\underline{a}^*} f_{a_1a_1}(x_1|a_1)dx_1$$

$$= \frac{\partial E(x_2|a_2)}{\partial a_2}\bigg|_{\underline{a}^*} - 2\lambda J - \mu_2^2 \int \frac{f_{a_2}^3}{f^2}\bigg|_{\underline{a}^*} dx_2$$

$$+ 2\mu_2^2 \int \frac{f_{a_2}}{f}\bigg|_{\underline{a}^*} f_{a_2a_2}(x_2|a_2)dx_2 \ . \tag{A1.7}$$

For the exponential family, using the results in Table II and equations (A1.2) through (A1.4), (A1.7) can be written as

$$B''(z(a_1^*))z'(a_1^*) - \mu_1^2[(z'(a_1^*))^3\ B'''(z(a_1^*)) - 2(z'(a_1^*))^3\ B'''(z(a_1^*))$$
$$- 2z'(a_1^*)z''(a_1^*)B''(z(a_1^*))]$$
$$= B''(z(a_2^*)z'(a_2^*) - \mu_2^2[(z'(a_2^*))^3\ B'''(z(a_2^*)) - 2(z'(a_2^*))^3\ B'''(z(a_2^*))$$
$$- 2z'(a_2^*)\ z''(a_2^*)\ B''(z(a_2^*))]\ ,$$

or

$$B''(z(a_1^*))z'(a_1^*) + \mu_1^2[(z'(a_1^*))^3\ B'''(z(a_1^*))$$

$$+\ 2z'(a_1^*)z''(a_1^*)B''(z(a_1^*))]$$

$$= B''(z(a_2^*))z'(a_2^*) + \mu_2^2[(z'(a_2^*))^3\ B'''(z(a_2^*))$$

$$+\ 2z'(a_2^*)z''(a_2^*)\ B''(z(a_2^*))] \qquad \text{(A1.8)}$$

Equation (A1.6) can be written as (see equation (A1.2))

$$\mu_1(z'(a_1^*))^2\ B''(z(a_1^*)) = \mu_2(z'(a_2^*))^2\ B''(z(a_2^*))\ . \qquad \text{(A1.9)}$$

Appendix 2

Normal Distribution Calculations

This appendix contains calculations for the bivariate normal distribution, the only distribution with a convenient representation for dependent random variables.

Suppose $\underline{x} \sim N(\underline{\theta}(\underline{a}), \Sigma(\underline{a}))$, with $\Sigma = \begin{pmatrix} \sigma_1^2(\underline{a}) & \rho(\underline{a})\sigma_1(\underline{a})\sigma_2(\underline{a}) \\ \rho(\underline{a})\sigma_1(\underline{a})\sigma_2(\underline{a}) & \sigma_2^2(\underline{a}) \end{pmatrix}$

where $\underline{a} = (a_1, a_2)$ and $\underline{\theta}(\underline{a}) = (\theta_1(\underline{a}), \theta_2(\underline{a}))^T$. Then

$$f(x_1, x_2 | a_1, a_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp[-\frac{B}{2(1-\rho^2)}], \text{ where} \tag{A2.1}$$

$$B = (\frac{x_1-\theta_1}{\sigma_1})^2 - 2\rho(\frac{x_1-\theta_1}{\sigma_1})(\frac{x_2-\theta_2}{\sigma_2}) + (\frac{x_2-\theta_2}{\sigma_2})^2 .$$

Let D denote the argument in the exponentiation in (A2.1). The following quantity plays an important role in the determination of the optimal sharing rule.

$$\frac{\partial f/\partial a_1}{f} = \frac{\partial}{\partial a_1} \log f = \frac{\partial}{\partial a_i} [-\log 2\pi - \log \sigma_1 - \log \sigma_2$$

$$- \frac{1}{2}\log(1-\rho^2) + D]$$

$$= -\frac{\sigma_1^{(1)}}{\sigma_1} - \frac{\sigma_2^{(1)}}{\sigma_2} + \frac{\rho\rho^{(1)}}{1-\rho^2} + D^{(1)} ,$$

where the superscript (i) denotes differentiation with respect to a_i.

$$D^{(1)} = -\frac{1}{2}\left[\frac{2\rho\rho^{(1)}}{(1-\rho^2)^2} B + \frac{1}{1-\rho^2} B^{(1)}\right] \text{ and}$$

$$B^{(1)} = 2\left[\frac{x_1-\theta_1}{\sigma_1}\right]\left[\frac{-\theta_1^{(1)}\sigma_1-(x_1-\theta_1)\sigma_1^{(1)}}{\sigma_1^2}\right] - 2\rho^{(1)}\left[\frac{x_1-\theta_1}{\sigma_1}\right]\left[\frac{x_2-\theta_2}{\sigma_2}\right]$$

$$- 2\rho\left[\frac{-\theta_1^{(1)}\sigma_1-(x_1-\theta_1)\sigma_1^{(1)}}{\sigma_1^2}\right]\left[\frac{x_2-\theta_2}{\sigma_2}\right]$$

$$- 2\rho\left[\frac{x_1-\theta_1}{\sigma_1}\right]\left[\frac{-\theta_2^{(1)}\sigma_2-(x_2-\theta_2)\sigma_2^{(1)}}{\sigma_2^2}\right]$$

$$+ 2\left[\frac{x_2-\theta_2}{\sigma_2}\right]\left[\frac{-\theta_2^{(1)}\sigma_2-(x_2-\theta_2)\sigma_2^{(1)}}{\sigma_2^2}\right] .$$

<u>Case (a)</u>: $\underline{\theta}(\underline{a}) = (\theta_1(\underline{a}),\ \theta_2(\underline{a}))^T$.

$$\frac{f_{a_1}}{f} = -\frac{\sigma_1^{(1)}}{\sigma_1} - \frac{\sigma_2^{(1)}}{\sigma_2} + \frac{\rho\rho^{(1)}}{1-\rho^2} - \frac{\rho\rho^{(1)}}{(1-\rho^2)^2} B - \frac{1}{2(1-\rho^2)} B^{(1)} .$$

Since f is symmetric, simply replace the (1)'s with (2)'s to get

$$\frac{f_{a_2}}{f} = -\frac{\sigma_1^{(2)}}{\sigma_1} - \frac{\sigma_2^{(2)}}{\sigma_2} + \frac{\rho\rho^{(2)}}{1-\rho^2} - \frac{\rho\rho^{(2)}}{(1-\rho^2)^2} B - \frac{1}{2(1-\rho^2)} B^{(2)} .$$

The optimal sharing rule is <u>not</u> separable in x_1 and x_2 and is <u>not</u> a commission scheme.

<u>Case (b)</u>: $\rho \equiv \text{constant} \neq 0$.

The optimal sharing rule will clearly not be a commission scheme. Of course, if $\rho \equiv 0$, the optimal sharing rule will be separable in x_1 and x_2 but will <u>not</u> be a commission (linear) scheme.

Case (c): $\rho \equiv$ constant and $\sigma_i^2 \equiv$ constant.

$$\frac{f_{a_1}}{f} = \frac{1}{1-\rho^2}\left[\frac{\theta_1^{(1)}(x_1-\theta_1)}{\sigma_1^2} - \frac{\rho}{\sigma_1\sigma_2}\left(\theta_1^{(1)}(x_2-\theta_2) + \theta_2^{(1)}(x_1-\theta_1)\right) + \frac{\theta_2^{(1)}(x_2-\theta_2)}{\sigma_2^2}\right].$$

$$\frac{f_{a_2}}{f} = \frac{1}{1-\rho^2}\left[\frac{\theta_2^{(2)}(x_2-\theta_2)}{\sigma_2^2} - \frac{\rho}{\sigma_1\sigma_2}\left(\theta_2^{(2)}(x_1-\theta_1) + \theta_1^{(2)}(x_2-\theta_2)\right) + \frac{\theta_1^{(2)}(x_1-\theta_1)}{\sigma_2^2}\right].$$

The optimal sharing rule will be a commission scheme with the coefficient of x_1 equal to

$$\frac{1}{1-\rho^2}\left[\mu_1\left(\frac{\theta_1^{(1)}}{\sigma_1^2} - \frac{\rho\theta_2^{(1)}}{\sigma_1\sigma_2}\right) + \mu_2\left(\frac{\theta_1^{(2)}}{\sigma_1^2} - \frac{\rho\theta_2^{(2)}}{\sigma_1\sigma_2}\right)\right].$$

The coefficient of x_2 is

$$\frac{1}{1-\rho^2}\left[\mu_1\left(\frac{\theta_2^{(1)}}{\sigma_2^2} - \frac{\rho\theta_1^{(1)}}{\sigma_1\sigma_2}\right) + \mu_2\left(\frac{\theta_2^{(2)}}{\sigma_2^2} - \frac{\rho\theta_1^{(2)}}{\sigma_1\sigma_2}\right)\right].$$

Lemma 2A.1 characterizes some properties of the optimal second best solution for particular utility functions when the distribution of the outcomes is bivariate normal. The calculations in the proof will be useful in the proofs of propositions in Section 3.5.

Lemma 2A.1. Suppose $(x_1, x_2) \sim N\left(\begin{pmatrix} k_1 a_1 \\ k_2 a_2 \end{pmatrix}, \Sigma\right)$,

where $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$. Suppose further that the principal is risk neutral, $U(s) = 2\sqrt{s}$, and $V(\underline{a}) = V(a_1 + a_2)$. Then assuming the interior characterization of the optimal sharing rule, $s(x_1,x_2) = (\lambda + \Sigma\mu_j f_{a_j}/f\,)^2$, is valid for almost every (x_1,x_2), the following results hold:

(1) $a_1^* > 0$, $a_2^* > 0$, $k_1 = k_2$, and $\sigma_1 = \sigma_2$ imply that $\mu_1 = \mu_2$.

(2) $k_1 \neq k_2$ implies that the optimal solution is a boundary solution, i.e., $a_1^* = 0$ or $a_2^* = 0$.

<u>Proof of Lemma 2A.1.</u> In this case,

$$\frac{f_{a_1}}{f} = [\,\frac{k_1(x_1-k_1a_1)}{\sigma_1^2} - \frac{\rho k_1(x_2-k_2a_2)}{\sigma_1\sigma_2}\,]/(1-\rho^2)\ ,$$

$$\frac{f_{a_2}}{f} = [\,\frac{k_2(x_2-k_2a_2)}{\sigma_2^2} - \frac{\rho k_2(x_1-k_1a_1)}{\sigma_1\sigma_2}\,]/(1-\rho^2)\ ,$$

and $s(x_1,x_2) = (\lambda + \sum_{j=1}^{2} \mu_j f_{a_j}/f\Big|_{\underline{a}^*}\,)^2$.

Let $$C_1 = \frac{\mu_1 k_1}{\sigma_1(1-\rho^2)} - \frac{\mu_2 \rho k_2}{\sigma_2(1-\rho^2)}$$ and

$$C_2 = \frac{\mu_2 k_2}{\sigma_2(1-\rho^2)} - \frac{\mu_1 \rho k_1}{\sigma_1(1-\rho^2)}\ .$$ The principal's expected return is

$$EW = E(x_1 + x_2 - s(x_1,x_2))$$

$$= k_1a_1 + k_2a_2 - E(\lambda + \frac{x_1-k_1a_1^*}{\sigma_1}\,C_1 + \frac{x_2-k_2a_2^*}{\sigma_2}\,C_2)^2$$

$$= k_1a_1 + k_2a_2 - E(\lambda + \frac{k_1a_1 - k_1a_1^*}{\sigma_1}\,C_1 + \frac{k_2a_2 - k_2a_2^*}{\sigma_2}\,C_2$$

$$+ \frac{x_1-k_1a_1}{\sigma_1}\,C_1 + \frac{x_2-k_2a_2}{\sigma_2}\,C_2)^2\ .$$

Letting $A_i = \frac{k_i a_i - k_i a_i^*}{\sigma_i} C_i$ and $y_i = \frac{x_i - k_i a_i}{\sigma_i}$, the principal's expected return can be written as

$$k_1 a_1 + k_2 a_2 - (\lambda + A_1 + A_2)^2 - E(\frac{x_1 - k_1 a_1}{\sigma_1} C_1)^2$$

$$- E(\frac{x_2 - k_2 a_2}{\sigma_2} C_2)^2 - 2C_1 C_2 E[(\frac{x_1 - k_1 a_1}{\sigma_1})(\frac{x_2 - k_2 a_2}{\sigma_2})]$$

$$= k_1 a_1 + k_2 a_2 - (\lambda + A_1 + A_2)^2 - C_1^2 \mathrm{Var}(y_1)$$

$$- C_2^2 \mathrm{Var}(y_2) - 2C_1 C_2 \mathrm{Cov}(y_1, y_2)$$

$$= k_1 a_1 + k_2 a_2 - (\lambda + A_1 + A_2)^2 - C_1^2 - C_2^2 - 2C_1 C_2 \rho . \qquad \text{(A2.2)}$$

$$\frac{\partial EW}{\partial a_i} = k_i - 2(\lambda + A_1 + A_2) \frac{k_i C_i}{\sigma_i}, \quad i=1,2, \text{ and}$$

$$\left.\frac{\partial EW}{\partial a_i}\right|_{a^*} = k_i - \frac{2\lambda k_i}{\sigma_i(1-\rho^2)} [\frac{\mu_i k_i}{\sigma_i} - \frac{\mu_j \rho k_j}{\sigma_j}] , \quad i,j=1,2, \; i \neq j.$$

If x_1 and x_2 are independent, then $\rho = 0$ and

$$\left.\frac{\partial EW}{\partial a_i}\right|_{\underline{a}^*} = k_i - 2\lambda k_i^{\,2} \mu_i / \sigma_i^2 .$$

Letting $a = a_1 + a_2$, the agent's expected utility is

$$EU = 2\, E(\lambda + \Sigma\, \mu_j f_{a_j}/f \big|_{\underline{a}^*}) - V(a)$$

$$= 2[\lambda + \frac{k_1 a_1 - k_1 a_1^*}{1 - \rho^2} (\frac{\mu_1 k_1}{\sigma_1^2} - \frac{\mu_2 k_2 \rho}{\sigma_1 \sigma_2})$$

$$+ \frac{k_2 a_2 - k_2 a_2^*}{1 - \rho^2} (\frac{\mu_2 k_2}{\sigma_2^2} - \frac{\mu_1 k_1 \rho}{\sigma_1 \sigma_2})] - V(a) .$$

$$\frac{\partial EU}{\partial a_i} = \frac{2k_i}{1 - \rho^2} (\frac{\mu_i k_i}{\sigma_i^2} - \frac{\mu_j k_j \rho}{\sigma_i \sigma_j}) - V'(a), \qquad i,j=1,2, \qquad i \neq j.$$

$$= \frac{2k_i C_i}{\sigma_i} - V'(a), \qquad i=1,2. \tag{A2.3}$$

The first order conditions require that (A2.3) is zero for i=1,2. Therefore,

$$\frac{k_1 C_1}{\sigma_1} = \frac{k_2 C_2}{\sigma_2} . \tag{A2.4}$$

If $k_1 = k_2$ and $\sigma_1 = \sigma_2$, then (A2.4) implies that $C_1 = C_2$, which in turn implies that $\mu_1 = \mu_2$ (assuming $\rho \neq \pm 1$). This establishes result (1) of Lemma 2A.1. Note that if $\rho = 0$, then setting (A2.3) equal to zero shows that $\mu_1 > 0$ and $\mu_2 > 0$, since V'(a) is assumed to be positive.

The Hamiltonian is

$$H = EW + \lambda(EU-\bar{u}) + \sum_{j=1}^{2} \mu_j \; \partial EU/\partial a_j \; .$$

$$\frac{\partial H}{\partial a_i} = k_i - 2[\; \lambda + \sum_{j=1}^{2} (a_j - a_j^*) \frac{k_j C_j}{\sigma_j} \;] \frac{k_i C_i}{\sigma_i}$$

$$- V''(a) \sum_{j=1}^{2} \mu_j , \qquad i=1,2. \tag{A2.5}$$

Setting $\frac{\partial H}{\partial a_i} = 0$ for i=1,2, and letting P denote the quantity in (A2.4) yields

$$k_1 - 2P[\; \lambda + P(a_1+a_2 - a_1^*-a_2^*) \;]$$

$$= k_2 - 2P[\; \lambda + P(a_1+a_2 - a_1^*-a_2^*) \;] . \tag{A2.6}$$

It is impossible to satisfy equation (A2.6) unless $k_1 = k_2$, which establishes result (2) in Lemma 2A.1.

Q.E.D.

Appendix 3

Chapter 3 Proofs

Proof of Proposition 3.1.1.

The principal's problem is

Maximize $EW(x-s(\underline{x})) = \int W(x-s(\underline{x})f(\underline{x}|\underline{a})d\underline{x}$

subject to $EU(s(\underline{x})) - V(\underline{a}) = \bar{u}$.

The first order condition for $s^*(\underline{x})$ requires that

$$- W'(x-s^*(\underline{x}))f(\underline{x}|\underline{a}) + \lambda\, U'(s^*(\underline{x}))f(\underline{x}|\underline{a}) = 0,$$

or $$W'(x-s^*(\underline{x})) = \lambda\, U'(s^*(\underline{x})) \,.$$

This implies that

$$x - s^*(\underline{x}) = W'^{-1}(\lambda\, U'(s^*(\underline{x}))) \equiv T(s^*(\underline{x})), \text{ with } T'(s^*) > 0 \text{ since } \lambda > 0.$$

Therefore, $x = T(s^*(\underline{x})) + s^*(\underline{x}) \equiv Y(s^*(\underline{x}))$, with $Y'(s^*) = T'(s^*) + 1 > 0$.

Thus, $s^*(\underline{x}) = Y^{-1}(x)$.

Q.E.D.

Lemma 3A.1 below will be used in proving Proposition 3.2.1.

Lemma 3A.1: Suppose $f(\underline{x}|\underline{a}) = \prod_{i=1}^{n} f^i(x_i|a_i)$ and that the risk-averse agent's expected utility is pseudoconcave in $\underline{a}$.

Suppose further that $F_{a_i}(x_i|a_i) \leq 0$, with strict inequality for some x_i-values. Then for $i=1,\ldots,n$, if $\mu_i \leq 0$, $\left.\frac{\partial EW}{\partial a_i}\right|_{\underline{a}^*} > 0$.

Proof of Lemma 3A.1: The first order conditions are

$$(1)\quad \int W(x-s(\underline{x}))f_{a_i}(\underline{x}|\underline{a}^*)d\underline{x} + \sum_{j=1}^{n} \mu_j^* \left\{ \int U(s|\underline{x})f_{a_ia_j}(\cdot)d\underline{x} - V_{a_ia_j} \right\} = 0,$$

$i=1,\ldots,n$, and

$$(2)\quad \frac{W'(r(\underline{x}))}{U'(x-r(\underline{x}))} = \lambda^* + \sum_{j=1}^{n} \mu_j^* \frac{f_{a_j}(\underline{x}|\underline{a}^*)}{f(\underline{x}|\underline{a}^*)}$$

$$= \lambda^* + \sum_{j=1}^{n} \mu_j^* \frac{f_{a_j}^{j}(x_j|a_j^*)}{f^j(x_j|a_j^*)}$$

because of the independence assumption. Here, subscripts a_i and a_j on $f(\cdot)$ and $V(\cdot)$ denote partial differentiation with respect to a_i or a_j, respectively; λ^* and μ_j^*, $j=1,\ldots,n$, are the optimal values of the multipliers in the second best problem, and $r(\underline{x}) = x-s(\underline{x})$.

Suppose some $\mu_j < 0$. Without loss of generality, let $j=1$. Consider the following auxiliary problem:

$$\underset{s_\lambda}{\text{Max}} \int W(x-s_\lambda(\underline{x}))f(\underline{x}|\underline{a}^*)d\underline{x} + \lambda^*\left[\int U(s_\lambda(\underline{x}))f(\underline{x}|\underline{a}^*)d\underline{x} - V(\underline{a}^*)\right]$$

$$+ \sum_{j=2}^{n} \mu_j^* \left[\int U(s_\lambda(\underline{x}))f_{a_j}(\underline{x}|\underline{a}^*)d\underline{x} - V_{a_j}(\underline{a}^*)\right],$$

where $\underline{a}^*$, λ^* and $\mu_2^*,\ldots,\mu_n^*$ correspond to the optimal solution characterized by (1) and (2). Let $r_\lambda(\underline{x}) = x - s_\lambda(\underline{x})$.

For $\underline{x} \in X_{1+} = \{\ \underline{x}$ with x_1 such that $f_{a_1}^{1}(x_1|a_1^*) > 0\ \}$,

$$\frac{W'(r(\underline{x}))}{U'(x-r(\underline{x}))} = \lambda^* + \sum_{j=1}^{n} \mu_j^* \frac{f_{a_j}^{j}(x_j|a_j^*)}{f^j(x_j|a_j^*)} < \lambda^* + \sum_{j=2}^{n} \mu_j^* \frac{f_{a_j}^{j}(x_j|a_j^*)}{f^j(x_j|a_j^*)}$$

$$= \frac{W'(r_\lambda(\underline{x}))}{U'(x-r_\lambda(\underline{x}))}.$$

Note that $\frac{W'(r(\underline{x}))}{U'(x-r(\underline{x}))}$ is decreasing in $r(\underline{x})$ for every fixed $\underline{x}$. Further, $r_\lambda(\underline{x})$ is an increasing function of x_1, since

$$\frac{W''(r_\lambda(\underline{x}))\,\frac{\partial r_\lambda}{\partial x_1}\,U'(x-r_\lambda(\underline{x})) + W'(\cdot)\,U''(\cdot)(1-\frac{\partial r_\lambda}{\partial x_1})}{U'^2} = 0$$

implies that $\frac{\partial r_\lambda}{\partial x_1} = \frac{W'U''}{W''U' + W'U''} > 0$.

Now

$$\frac{W'(r(\underline{x}))}{U'(x-r(\underline{x}))} < \frac{W'(r_\lambda(\underline{x}))}{U'(x-r_\lambda(\underline{x}))} \text{ and } \frac{W'(r(\underline{x}))}{U'(x-r(\underline{x}))} \text{ decreasing in } r$$

implies that $r(\underline{x}) > r_\lambda(\underline{x})$, for all $\underline{x} \in X_{1+}$. Correspondingly, $r(\underline{x}) < r_\lambda(\underline{x})$ on $X_{1-} = \{\, \underline{x}$ with x_1 such that $f^1_{a_1}(x_1|a_1^*) < 0 \,\}$.

Therefore,

$$\int W(r(\underline{x}))f_{a_1}(\underline{x}|\underline{a}^*)d\underline{x} - \int W(r_\lambda(\underline{x}))f_{a_1}(\underline{x}|\underline{a}^*)d\underline{x}$$

$$= \int_{X_{1-}} [\, W(r(\underline{x})) - W(r_\lambda(\underline{x}))\,]\, f_{a_1}(\underline{x}|\underline{a}^*)d\underline{x}$$

$$+ \int_{X_{1+}} [\, W(r(\underline{x})) - W(r_\lambda(\underline{x}))\,]\, f_{a_1}(\cdot)d\underline{x} > 0.$$

It remains to show that $\int W(r_\lambda(\underline{x}))f_{a_1}(\underline{x}|\underline{a}^*)d\underline{x} > 0$. The left-hand side of the expression can be written as

$$\int_{X_1} [\, \int_{X_2} \cdots \int_{X_n} W(r_\lambda(\underline{x}))f^2(x_2|a_2^*)\cdots f^n(x_n|a_n^*)dx_2\cdots dx_n\,]\, f^1_{a_1}(x_1|a_1^*)dx_1$$

$= \int_{X_1} T(x_1)f^1_{a_1}(x_1|a_1^*)dx_1 > 0$, as in the one dimensional case, because of stochastic dominance and the fact that

$$T'(x_1) = \int W' \frac{\partial r_\lambda}{\partial x_1} f^2 \ldots f^n dx_2 \ldots dx_n > 0 .$$

Q.E.D.

<u>Proof of Proposition 3.2.1:</u> Let $A = \int W(x-s(\underline{x}))f(\underline{x}|\underline{a})d\underline{x}$ and

$$B = \int U(s(\underline{x}),\underline{a})f(\underline{x}|\underline{a})d\underline{x}.$$

Subscripts i and j on A and B will denote partial differentiation with respect to a_i or a_j, respectively. The first order conditions $\frac{\partial H}{\partial a_i} = 0$ for n=2 are

(1) $A_1 + \mu_1 B_{11} + \mu_2 B_{12} = 0$ and

(2) $A_2 + \mu_1 B_{21} + \mu_2 B_{22} = 0,$

where the functions are evaluated at the optimal $\underline{a}^*$ and with the optimal $s^*(\underline{x})$. In matrix notation,

$$A + B\ \mu = \underline{0},$$

where $A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$, $B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, and $\underline{0}$ $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

If B is strictly concave in $\underline{a}$, then $|B| \neq 0$ and B^{-1} exists. Therefore,

$$\mu = B^{-1}(-A) = \frac{1}{|B|} \begin{pmatrix} B_{22} & -B_{12} \\ -B_{21} & B_{11} \end{pmatrix} \begin{pmatrix} -A_1 \\ -A_2 \end{pmatrix} .$$

I.e.,

(3) $$\mu_1 = \frac{A_2 B_{12} - A_1\ B_{22}}{|B|}$$ and

(4) $$\mu_2 = \frac{A_1 B_{21} - A_2 B_{11}}{|B|} .$$

If B is strictly concave in $\underline{a}$, then $|B| > 0$ and $B_{ii} < 0$, i=1,2.

Now assume $\mu_1 \leq 0$ and $\mu_2 \leq 0$. Then by Lemma 3A.1, $A_1 > 0$ and $A_2 > 0$. From (3) and (4), we have

$A_2B_{12} - A_1B_{22} \leq 0$ and

$A_1B_{21} - A_2B_{11} \leq 0$.

These imply that $B_{12} < 0$ (note: $B_{12} = B_{21}$). But if $B_{12} < 0$, then (1) and (2) cannot be satisfied. Therefore, not both μ_1 and μ_2 can be nonpositive.

Q.E.D.

Lemma 3A.2 below deals with the problem of allocating effort to two tasks considered simultaneously.

<u>Lemma 3A.2.</u> (First Best, Additive Effort)

Suppose $E(x_i) = k_i a_i$, i=1,...,n.

(1) If $k_i = k$, for all i, then $k = \lambda V'(\Sigma a_i)$ implies that any nonnegative vector $\underline{a}$ such that Σa_i satisfies [1] below is Pareto optimal.

(2) If some $k_i \neq k_j$, a boundary solution results. That is, all the a_i's are zero except one. In the n=2 case with $k_1 > k_2$, $a_1^* > 0$ and $a_2^* = 0$.

<u>Proof of Lemma 3A.2.</u> The principal's problem is

$$\underset{s(\underline{x}),\ \underline{a}}{\text{Maximize}} \int (x-s(\underline{x}))\, g(\underline{x}|\underline{a})\, d\underline{x}$$

$$\text{subject to} \int [\, U(s(\underline{x})) - V(a)\,]\, g(\underline{x}|\underline{a})\, d\underline{x} \geq \bar{u}.$$

$$H = \int (x-s(\underline{x}))\, g(\underline{x}|\underline{a})\, d\underline{x} + \lambda \left\{ \int [\, U(s(\underline{x})) - V(a)\,]\, g(\underline{x}|\underline{a})\, d\underline{x} - \bar{u} \right\}.$$

$\frac{\partial H}{\partial s} = -g + \lambda U'g = 0$ implies that $U'(s(\underline{x})) = \frac{1}{\lambda}$, which implies that

$s(\underline{x}) = U'^{-1}(\frac{1}{\lambda}) = C.$

$$\frac{\partial H}{\partial a_i} = \int (x-s(\underline{x}))\; g_{a_i}(\underline{x}|\underline{a})\; d\underline{x} + \lambda \left\{ \int [\; U(s(\underline{x}))\; g_{a_i}(\cdot)\;]\; d\underline{x} - V'(\Sigma a_j) \right\} = 0 \;.$$

$$\frac{\partial E(x|\underline{a})}{\partial a_i} - 0 + \lambda(0 - V'(\Sigma a_j)) = 0 \quad \text{implies that} \quad k_i = \lambda V'(\Sigma a_j) \text{ for all } i. \qquad [1]$$

This establishes result (1) of Lemma 3A.2.

To establish result (2), recall that a_1^* and a_2^* are nonnegative by assumption. Let $s^*(\underline{x}) = C^*$, where $s^*(\underline{x})$ is the optimal sharing rule corresponding to the optimal choices a_1^* and a_2^*. Let (a_1', a_2'), where $a_1' > 0$ and $a_2' > 0$, be a feasible effort pair given C^*.

The agent's expected utility for any feasible (a_1,a_2) is

$$C^* - V(a_1 + a_2) = \bar{u} \;.$$

Since (a_1',a_2') is feasible, $C^* - V(a_1' + a_2') = \bar{u}$. Consider the pair $(a_1'',a_2'') = (a_1' + a_2', 0)$. This pair is also feasible, since $a_1'' + a_2'' = a_1' + a_2'$, and the principal is strictly better off with (a_1'',a_2'') since his expected return is

$$k_1 a_1'' + k_2 a_2'' - C^* = k_1(a_1' + a_2') - C^* > k_1 a_1' + k_2 a_2' - C^* \text{ if } k_1 > k_2.$$

Therefore, $a_2' > 0$ is not optimal, and hence the optimal effort pair is such that $a_1^* > 0$ and $a_2^* = 0$.

Q.E.D.

Proposition 3A.3 below compares the solutions to two one-task problems.

<u>Proposition 3A.3</u>. (First Best, Additive Effort)

Suppose $E(x_i) = k_i a_i$, i=1,2, and that $k_1 > k_2$. Consider the two separate problems where effort is devoted only to task i. Then

(1) $a_1^* > a_2^*$ if V is increasing and convex,

(2) $a_1^* > a_2^*$ implies that $s_1^* > s_2^*$ (i.e., the agent is paid more for exerting a_1^* at task 1 than for exerting a_2^* at task 2), and

(3) the principal is better off with $a_1^* > 0$ and $a_2^* = 0$ than with $a_1^* = 0$ and $a_2^* > 0$.

Proof of Proposition 3A.3.

The principal's problem if effort is devoted only to task i is

Problem i: $\underset{s_i(x), a_i}{\text{Maximize}} \int (x - s_i(x))\, g^i(x|a_i)\, dx$

subject to $\int U(s_i(x))\, g^i(x|a_i)\, dx - V(a_i) \geq \bar{u}$.

$$\frac{\partial H_i}{\partial s_i} = 0 \quad \text{implies that} \quad s_i(x) = U'^{-1}\left(\frac{1}{\lambda_i}\right) = C_i \quad \text{and} \qquad [1]$$

$$\frac{\partial H_i}{\partial a_i} = 0 \quad \text{implies that} \quad k_i = \lambda_i V'(a_i) = \frac{1}{\lambda_i} = \frac{V'(a_i)}{k_i} . \qquad [2]$$

Feasibility requires that $U\left(U'^{-1}\left(\frac{1}{\lambda_i}\right)\right) - V(a_i) = \bar{u}$,

which implies that $U'^{-1}\left(\frac{1}{\lambda_i}\right) = U'^{-1}\left[\bar{u} + V(a_i)\right]$. [3]

Equation [2] implies that

$$U'^{-1}\left(\frac{1}{\lambda_i}\right) = U'^{-1}\left(\frac{V'(a_i)}{k_i}\right) . \qquad [4]$$

Result (1). $k_1 > k_2$ implies that $a_1^* > a_2^*$ if V is increasing and convex.

Proof. Suppose $a_1^* < a_2^*$. Then

$U^{-1}(\bar{u} + V(a_1^*)) < U^{-1}(\bar{u} + V(a_2^*))$

(since U^{-1} and V are increasing)

implies that $U'^{-1}\left(\frac{V'(a_1^*)}{k_1}\right) < U'^{-1}\left(\frac{V'(a_2^*)}{k_2}\right)$ by [3] and [4], which implies that

$$\frac{V'(a_1^*)}{k_1} > \frac{V'(a_2^*)}{k_2} \quad \text{(since } U'^{-1} \text{ is decreasing), or}$$

$$\frac{k_2}{k_1} > \frac{V'(a_2^*)}{V'(a_1^*)} > 1 \quad \text{(since } V' \text{ is increasing and } a_1^* < a_2^*\text{), so that}$$

$$k_2 > k_1 .$$

Therefore, $k_1 > k_2$ implies that $a_1^* > a_2^*$.

Result (2). $a_1^* > a_2^*$ implies that $s_1^* > s_2^*$.

Proof. $a_1^* > a_2^*$ implies that $\bar{u} + V(a_1^*) > \bar{u} + V(a_2^*)$, which implies that

$$U^{-1}(\bar{u} + V(a_1^*)) > U^{-1}(\bar{u} + V(a_2^*))$$

(since U^{-1} is increasing), so that

$$U'^{-1}\left(\frac{1}{\lambda_1}\right) > U'^{-1}\left(\frac{1}{\lambda_2}\right) \quad \text{by [3]. Therefore,}$$

$$s_1^* > s_2^* \quad \text{by [1] .}$$

Remark: $a_1^* > a_2^*$ also implies that $\lambda_1 > \lambda_2$.

Result (3). If $k_1 > k_2$, the principal is better off with $a_1^* > 0$ and $a_2^* = 0$ than with $a_1^* = 0$ and $a_2^* > 0$.

Proof. It is necessary to show that

$$\int (x - s_1^*)\, g^1(x|a_1^*)\, dx > \int (x - s_2^*)\, g^2(x|a_2^*)\, dx,$$

that is, $k_1 a_1^* - C_1 > k_2 a_2^* - C_2$. [5]

Note that (a_2^*, s_2^*) is feasible for Problem 1:

$$\int U(s_2^*)\, g^1(x|a_2^*)\, dx - V(a_2^*) = U(C_2) - V(a_2^*) = \bar{u} .$$

Therefore, $k_1 a_1^* - s_1^* > k_1 a_2^* - s_2^*$ because of feasibility of (a_2^*, s_2^*) for Problem 1 and optimality of (a_1^*, s_1^*) for Problem 1. Furthermore, $k_1 a_2^* - s_2^* > k_2 a_2^* - s_2^*$ because $k_1 > k_2 > 0$, and hence [5] holds.

Q.E.D.

Proof of Proposition 3.5.1: In Lemma 2A.1 of Appendix 2, it was shown that if $x_i \sim N(ka_i, \sigma^2)$ and an interior solution $(a_1^* > 0,\ a_2^* > 0)$ is optimal, then it must be that $\mu_1 = \mu_2$. Let $\mu = \mu_1$, $\underline{a}^* = (a_1^*, a_2^*)$, and $a^* = a_1^* + a_2^*$. This interior solution satisfies the Nash conditions

$$2 \frac{\partial}{\partial a_i} \int \left(\lambda + \mu \sum_j \left. \frac{f_{a_j}(x_j|a_j)}{f(x_j|a_j)} \right|_{\underline{a}^*} \right) f(x_1|a_1) f(x_2|a_2) d\underline{x}$$

$$- V'(a_1 + a_2) = 0, \qquad i=1,2.$$

The condition for i=1 is

$$2\mu \left\{ \frac{\partial}{\partial a_1} \int \frac{f_{a_1}(x_1|a_1^*)}{f(x_1|a_1^*)} \cdot f(x_1|a_1) dx_1 \right.$$

$$\left. + \frac{\partial}{\partial a_1} \int \frac{f_{a_2}(x_2|a_2^*)}{f(x_2|a_2^*)} \cdot f(x_1|a_1) f(x_2|a_2) d\underline{x} \right\} - V'(a_1 + a_2) = 0,$$

or $2\mu \dfrac{\partial}{\partial a_1} \int (kx_1 - k^2 a_1^*) \cdot f(x_1|a_1) dx_1 - V'(a_1 + a_2) = 0$, i.e., $2\mu k^2 - V'(a^*) = 0$, which would also result from the i=2 condition. Hence, there is really only one Nash condition. The principal's expected utility is

$$\int \left(x_1 + x_2 - \left(\lambda + \mu \sum_j \frac{f_{a_j}(x_j|a_j^*)}{f(x_j|a_j^*)} \right)^2 \right) f(x_1|a_1^*)f(x_2|a_2^*)dx_1dx_2$$

$= ka^* - \lambda^2 - 2\mu^2k^2$ (see equation (A2.2) in Appendix 2).

The agent's expected utility if effort a* is exerted is

$$2 \int \left(\lambda + \mu \sum_j \frac{f_{a_j}(x_j|a_j^*)}{f(x_j|a_j^*)} \right) f(x_1|a_1^*)f(x_2|a_2^*)dx_1dx_2 - V(a^*) = \bar{u} ,$$

which implies that $2\lambda - V(a^*) = \bar{u}$.

Now suppose that a_2=0, the minimum effort, and that x_2 is ignored for compensation purposes. Consider $s(x) = \left(\lambda + \mu \left.\frac{f_a(x|a)}{f(x|a)}\right|_{a^*} \right)^2$ where λ, μ, and a* are the same as in the interior solution above. The Nash condition is now

$$2 \frac{\partial}{\partial a} \int \left(\lambda + \mu \left.\frac{f_a(x|a)}{f(x|a)}\right|_{a^*} \right) f(x|a)dx - V'(a) = 0 ,$$

or $\quad 2\mu \frac{\partial}{\partial a} \int (kx - k^2a^*)f(x|a)dx - V'(a) = 0 ,$

that is, $2\mu k^2 - V'(a) = 0$, which is satisfied at a=a*.

The principal's expected utility if a=a* is

$$\int \left(x - \left(\lambda + \mu \left.\frac{f_a(x|a)}{f(x|a)}\right|_{a^*} \right)^2 \right) f(x|a^*)dx$$

$= ka^* - \lambda^2 - \mu^2k^2$

$> ka^* - \lambda^2 - 2\mu^2k^2$, which is the principal's expected utility with an interior solution. Since the agent's expected utility is unaffected, the principal is strictly better off, and the Nash condition holds, a boundary solution is optimal.

Q.E.D.

Proof of Proposition 3.5.2: The Hamiltonian for the two-task problem is

$$H = E(x_1 + x_2 - s^*(\underline{x})) + \lambda [EU(s^*(\underline{x})) - V(a_1 + a_2) - \bar{u}]$$
$$+ \mu_1 \frac{\partial}{\partial a_1} [EU(s^*(\underline{x})) - V(a_1 + a_2)]$$
$$+ \mu_2 \frac{\partial}{\partial a_2} [EU(s^*(\underline{x})) - V(a_1 + a_2)] .$$

The first order conditions are

$$\left.\frac{\partial H}{\partial a_i}\right|_{a^*} = 0, \qquad [1]$$

$$\left.\frac{\partial H}{\partial s}\right|_{a^*} = 0 \text{ pointwise}, \qquad [2]$$

$$\left.\frac{\partial}{\partial a_i} [EU(s^*(\underline{x})) - V(a_1 + a_2)]\right|_{a^*} = 0 , \qquad [3]$$

and $$EU(s^*(\underline{x})) - V(a_1^* + a_2^*) = \bar{u} . \qquad [4]$$

As before, [2] implies that

$$\frac{1}{U'(s^*(\underline{x}))} = \lambda + \sum_{j=1}^{2} \mu_j \left.\frac{f_{a_j}(x_j|a_j)}{f(x_j|a_j)}\right|_{a^*} .$$

$$\left.\frac{\partial H}{\partial a_1}\right|_{a^*} = \left.\frac{\partial}{\partial a_1} \int (x_1 + x_2 - s^*(\underline{x}))f(x_1|a_1)f(x_2|a_2)d\underline{x}\right|_{a^*} + 0$$

$$+ \mu_1 \left.\frac{\partial^2}{\partial a_1^2} [EU(s^*(\underline{x})) - V(\cdot)]\right|_{a^*}$$

$$+ \mu_2 \frac{\partial^2}{\partial a_1 \partial a_2} [\ EU(s^*(\underline{x})) - V(\cdot)\ ]\Big|_{a^*} = 0 .$$

$$\frac{\partial H}{\partial a_2}\Big|_{a^*} = \frac{\partial}{\partial a_2} \int (x_1 + x_2 - s^*(\underline{x}))f(x_1|a_1)f(x_2|a_2)d\underline{x}\Big|_{a^*} + 0$$

$$+ \mu_1 \frac{\partial^2}{\partial a_1 \partial a_2} [\ EU(s^*(\underline{x})) - V(\cdot)\ ]\Big|_{a^*}$$

$$+ \mu_2 \frac{\partial^2}{\partial a_2^2} [\ EU(s^*(\underline{x})) - V(\cdot)\ ]\Big|_{a^*} = 0 .$$

These imply that

$$\frac{\partial H}{\partial a_1}\Big|_{a^*} = \frac{\partial H}{\partial a_2}\Big|_{a^*} . \qquad [5]$$

Similarly, it is necessary that

$$\frac{\partial}{\partial a_1} \int U(s^*(\underline{x}))f(x_1|a_1)f(x_2|a_2)d\underline{x}\Big|_{a^*} = V'(a_1{}^* + a_2{}^*)$$

$$= \frac{\partial}{\partial a_2} \int U(s^*(\underline{x}))f(x_1|a_1)f(x_2|a_2)d\underline{x}\Big|_{a^*} . \qquad [6]$$

It is clear that ($a_1{}^* = a_2{}^*$ and $\mu_1{}^* = \mu_2{}^*$) constitute a solution to conditions [5] and [6]. Therefore, if a unique interior solution is optimal, then it has $a_1{}^* = a_2{}^*$ and $\mu_1{}^* = \mu_2{}^*$. The particular values of $a_1{}^*$, $\mu_1{}^*$, and λ are determined from conditions [1], [3], and [4].

Q.E.D.

Proof of Proposition 3.5.3: Consider first the situation where $a_2 = 0$ and the agent's compensation is based only on x_1. Dropping the subscript for convenience, the optimal sharing rule is

$$t(x) = [\ \lambda_0 + \mu_0 \left.\frac{f_a(x|a)}{f(x|a)}\right|_{a*}\ ]^2$$

$$= [\ \lambda_0 + \mu_0 z'(a*)(x - E(x|a*))\ ]^2,$$

where $\mu_0 > 0$ (Holmstrom, 1979). Recall that $E(x|a) = B'(z(a))$. The principal's expected return is

$$\int (x - t(x))f(x|a*)dx$$

$$= B'(z(a*)) - \lambda_0^2 - \mu_0^2(z'(a*))^2 \int (x-E(x|a*))^2 f(x|a*)dx$$

$$= B'(z(a*)) - \lambda_0^2 - \mu_0^2(z'(a*))^2\ B''(z(a*))\ , \tag{1}$$

since $\quad Var(x|a*) = B''(z(a*))$.

The agent's expected utility is

$$2 \int (\ \lambda_0 + \mu_0 \frac{f_a(x|a*)}{f(x|a*)}\ )\ f(x|a*)dx - V(a*) = \bar{u}\ ,$$

which implies that

$$2\lambda_0 - V(a*) = \bar{u}\ .$$

The Nash condition is

$$2 \int (\ \lambda_0 + \mu_0 \frac{f_a(x|a*)}{f(x|a*)}\ )\ f_a(x|a*)dx - V'(a*) = 0,$$

$$\text{or } 2\mu_0 \int \frac{f_a^2\ (x|a*)}{f(x|a*)}\, dx - V'(a*) = 0\ .$$

Now consider the two-task situation, where $f(x_1|a) = f(x_2|a)$ if $x_1 = x_2$.

Let

$$s(x_1,x_2) = \left(\lambda_0 + \mu \sum_{i=1}^{2} \frac{f_{a_i}(x_i|a_i)}{f(x_i|a_i)}\bigg|_{\underline{a}^*} \right)^2 ,$$

where $\underline{a}' = (a_1', a_2')$ and $a_1' = a_2' = \frac{a^*}{2}$.

The Nash conditions are now

$$2 \int \left(\lambda_0 + \mu \sum_{i=1}^{2} \frac{f_{a_i}(x_i|a_i')}{f(x_i|a_i')} \right) f_{a_1}(x_1|a_1)f(x_2|a_2)dx_1dx_2$$

$$- V'(a_1 + a_2) = 0 \quad \text{and}$$

$$2 \int \left(\lambda_0 + \mu \sum_{i=1}^{2} \frac{f_{a_i}(x_i|a_i')}{f(x_i|a_i')} \right) f(x_1|a_1)f_{a_2}(x_2|a_2)dx_1dx_2$$

$$- V'(a_1 + a_2) = 0 .$$

When evaluated at $\underline{a}'$, the Nash conditions reduce to

$$2 \mu \int \frac{f^2_{a_1}(x_1|a_1')}{f(x_1|a_1')} dx_1 = V'(a^*) = 2 \mu \int \frac{f^2_{a_2}(x_2|a_2')}{f(x_2|a_2')} dx_2 .$$

The Nash conditions will thus hold at $\underline{a}'$ if

$$\mu \int \frac{f^2_{a_1}(x_1|a_1')}{f(x_1|a_1')} dx_1 = \mu_0 \int \frac{f^2_a(x|a^*)}{f(x|a^*)} dx ,$$

that is, if

$$C \equiv \mu(z'(a_1'))^2 B''(z(a_1')) = \mu_0(z'(a^*))^2 B''(z(a^*)) \qquad [2]$$

(see equation (A1.9) in Appendix 1).

Equation [2] is true if

$$\mu = \frac{(z'(a^*))^2 \; B''(z(a^*))}{(z'(\frac{a^*}{2}))^2 \; B''(z(\frac{a^*}{2}))} \cdot \mu_0 \equiv \frac{z'(a^*)M'(a^*)}{z'(\frac{a^*}{2})M'(\frac{a^*}{2})} \cdot \mu_0 \; . \qquad [3]$$

The principal's expected return is

$$\int (x_1 + x_2 - s(x_1,x_2))f(x_1|a_1')f(x_2|a_2')dx_1dx_2$$

$$= 2B'(z(\tfrac{a^*}{2})) - \lambda_0{}^2 - 2\mu^2(z'(\tfrac{a^*}{2}))^2 \; B''(z(\tfrac{a^*}{2})) \; . \qquad [4]$$

Since M(a) is concave, $M(a^*) \leq 2M(\frac{a^*}{2})$. (Proof: $\frac{1}{2}M(0) + \frac{1}{2}M(a^*)$

$\leq M(\frac{a^*}{2})$ because $M(\cdot)$ is concave. If $M(0) \geq 0$, then $\frac{1}{2}M(a^*) \leq M(\frac{a^*}{2})$).

That is, $B'(z(a^*)) \leq 2B'(z(\frac{a^*}{2}))$. [5]

Suppose that $\dfrac{z'(a^*) \; M'(a^*)}{z'(\frac{a^*}{2}) \; M'(\frac{a^*}{2})} < \frac{1}{2}$. [6]

The difference between [4] and [1] is

$$2B'(z(\tfrac{a^*}{2})) - B'(z(a^*)) - 2\mu C + \mu_0 C > 0$$

because [5] holds and because [3] and [6] imply that

$$\mu_0 - 2\mu = \mu_0 \left[1 - \frac{2z'(a^*)M'(a^*)}{z'(\frac{a^*}{2})M'(\frac{a^*}{2})} \right] > 0.$$

If M(a) is strictly concave, then the inequality in [5] becomes a strict inequality, and hence the strict inequality in [6] can be relaxed to be a nonstrict inequality ($\leq$).

Finally, the agent's expected utility in the two-task situation described above is still $2\lambda_0 - V(a^*) = \bar{u}$. Since the principal is better off with an interior solution which satisfies the Nash conditions, a boundary solution is not optimal.

Q.E.D.

Proof of Corollary 3.5.4: If $M(a) = ka$, then (3.5.5) reduces to

$z'(a^*)/z'(\frac{a^*}{2}) < \frac{1}{2}$.

(i) For the exponential distribution with mean ka,

$z(a) = \frac{-1}{ka}$ (see Table II in Appendix 1), and hence

$$z'(a)/z'(\tfrac{a}{2}) = \frac{1}{ka^2} \Big/ \frac{4}{ka^2} = \tfrac{1}{4} < \tfrac{1}{2} .$$

(ii) For the gamma distribution with mean ka,

$z(a) = \frac{-n}{ka}$ (see Table II in Appendix 1), and hence

$$z'(a)/z'(\tfrac{a}{2}) = \frac{1}{ka^2} \Big/ \frac{4}{ka^2} = \tfrac{1}{4} < \tfrac{1}{2} .$$

(iii) For the normal distribution with mean ka and unit variance,

$z(a) = ka$ (see Table II in Appendix 1). Therefore,

$z'(a)/z'(\frac{a}{2}) = 1 > \frac{1}{2}$.

(iv) For the Poisson distribution with mean ka,

$z(a) = \ln ka = \ln k + \ln a$ (see Table II in Appendix 1).

$z'(a)/z'(\frac{a}{2}) = \frac{1}{a} \Big/ \frac{2}{a} = \frac{1}{2}$.

Q.E.D.

Proof of Proposition 3.5.5: In this case, $B'(z(a)) = ka$, $B''(z(a))z'(a) = k$, and $B'''(z(a))(z'(a))(z'(a))^2 + B''(z(a))z''(a) = 0$.

Equation (A1.8) reduces to

$$k + \mu_1^2\, z'(a_1^*)z''(a_1^*)B''(z(a_1^*)) = k + \mu_2^2\, z'(a_2^*)z''(a_2^*)B''(z(a_2^*)) ,$$

which implies that

$$\mu_1^2\, z''(a_1^*) = \mu_2^2\, z''(a_2^*) . \qquad [1]$$

Equation (A1.9) reduces to

$$\mu_1 z'(a_1^*) = \mu_2 z'(a_2^*) . \qquad [2]$$

[1] and [2] together imply that

$$\frac{z''(a_1^*)}{(z'(a_1^*))^2} = \frac{z''(a_2^*)}{(z'(a_2^*))^2} . \qquad [3]$$

Let $v(a) = z''(a)/(z'(a))^2$. If $v(a)$ is strictly monotone, then [3] implies that $a_1^* = a_2^*$. This in turn implies, from [1] or [2], that $\mu_1^* = \mu_2^*$.

Q.E.D.

Examples

1) Exponential: $z(a) = \frac{-1}{ka}$, $z'(a) = \frac{1}{ka^2}$, $z''(a) = \frac{-2}{ka^3}$, and

$$v_e(a) = \frac{\frac{-2}{ka^3}}{\frac{1}{k^2a^4}} = -2ka .$$

$v_e(a)$ is strictly decreasing in a, so $a_1^* = a_2^*$ and $\mu_1^* = \mu_2^*$.

2) Gamma: $z(a) = \frac{-n}{ka}$, so $v_g(a)$ is a constant multiple of $v_e(a)$. Therefore, $a_1^* = a_2^*$ and $\mu_1^* = \mu_2^*$.

3) Normal with unit variance: $z(a) = ka$, $z'(a) = k$, $z''(a) = 0$, and $v_n(a) \equiv 0$. Recall that in this case, a boundary solution is optimal. If an interior solution is required, [2] indicates that the multipliers μ_1 and μ_2 would have to be equal.

4) Poisson: $z(a) = \ln ka$, $z'(a) = \frac{1}{a}$, $z''(a) = -\frac{1}{a^2}$, and

$$v_p = \frac{-\frac{1}{a^2}}{\frac{1}{a^2}} \equiv -1 .$$ Therefore, the solution is not unique and we cannot say whether $\mu_1 = \mu_2$.

Proof of Proposition 3.5.6: Equation (A1.6) can be written as

$$\mu_1 I(a_1^*) = \mu_2 I(a_2^*). \qquad [1]$$

Note that $I'(a) \equiv \frac{d}{da} \int \frac{f_a^2(x|a)}{f(x|a)} dx = \int \left(\frac{2f_a f_{aa}}{f} - \frac{f_a^3}{f^2} \right) dx,$

and hence equation (A1.7) can be written as

$$\mu_1^2 \, I'(a_1^*) = \mu_2^2 \, I'(a_2^*) \ . \qquad [2]$$

Equations [1] and [2] together imply that

$$\frac{I'(a_1^*)}{I^2(a_1^*)} = \frac{I'(a_2^*)}{I^2(a_2^*)} \ .$$

Therefore, if $I'(a)/I^2(a)$ is strictly monotonic, then $a_1^* = a_2^*$, which implies that $\mu_1^* = \mu_2^*$ (equation [1]).

Q.E.D.

Proof of Corollary 3.5.7. For cases (i) - (iii), an interior solution is optimal if (3.5.6) holds.

(i) $z(a) = M(a)$, and hence (3.5.6) requires that

$$\left(\frac{\alpha a^{\alpha-1}}{\alpha(\frac{a}{2})^{\alpha-1}} \right)^2 < 1/2 , \quad \text{or} \quad 2^{2\alpha-2} < 1/2 ,$$

which is satisfied if $0 < \alpha < 1/2$.

(ii) $z(a) = \frac{-1}{M(a)}$, and hence (3.5.6) requires that

$$\left(\frac{M'(a)}{M(a)} \div \frac{M'(\frac{a}{2})}{M(\frac{a}{2})} \right)^2 < 1/2 , \quad \text{i.e.,} \quad \left(\frac{M'(a)}{M'(\frac{a}{2})} \cdot \frac{M(\frac{a}{2})}{M(a)} \right)^2 < 1/2 ,$$

or $2^{2\alpha-2} \cdot (\frac{\frac{a}{2}}{a})^{2\alpha} < 1/2$, which is satisfied when $0 < \alpha < 1$, since

$2^{2\alpha-2-2\alpha} = \frac{1}{4} < \frac{1}{2}$.

(iii) $z(a) = \ln M(a)$, and hence $z'(a) = \frac{M'(a)}{M(a)} = \frac{\alpha}{a}$.

Equation (3.5.6) requires that

$$\frac{\frac{\alpha}{a} \cdot \alpha a^{\alpha-1}}{\frac{2\alpha}{a} \cdot \alpha(\frac{a}{2})^{\alpha-1}} < 1/2 , \quad \text{i.e.,}\ 1/2 \cdot 2^{\alpha-1} < 1/2 ,$$

which is satisfied if $0 < \alpha < 1$.

Q.E.D.

<u>Proof of Proposition 3.5.8.</u> In this case, $B'(z(a)) = ka$, $B''(z)z'(a)=k$, and $B'''(z)z'(a)^2 + B''(z)z''(a) = 0$. Equation (A1.9) says that

$$\mu_1 k_1 z_1'(a_1^*) = \mu_2 k_2 z_2'(a_2^*) , \qquad [1]$$

and equation (A1.8) says that

$$k_1 + \mu_1^2 k_1 z_1''(a_1^*) = k_2 + \mu_2^2 k_2 z_2''(a_2^*) . \qquad [2]$$

(i) $z(a) = -\frac{1}{ka}$, $z'(a) = \frac{1}{ka^2}$, and $z''(a) = -\frac{2}{ka^3}$.

Equations [1] and [2] become

$$\frac{\mu_1^*}{a_1^{*2}} = \frac{\mu_2^*}{a_2^{*2}} \equiv R_e \qquad [3]$$

and $k_1 + \mu_1^2 \cdot \frac{-2}{a_1^{*3}} = k_2 + \mu_2^2 \cdot \frac{-2}{a_2^{*3}}$, which together imply that

$$k_1 - 2R_e^2\, a_1{}^* = k_2 - 2R_e^2\, a_2{}^* ,$$

or $$2R_e^2\, (a_1{}^* - a_2{}^*) = k_1 - k_2 > 0$$

since $k_1 > k_2$. Therefore, $a_1{}^* > a_2{}^*$, and hence, $\mu_1{}^* > \mu_2{}^*$ (from [3]).

(ii) $z(a) = -\frac{n}{ka}$, $z'(a) = \frac{n}{ka^2}$, and $z''(a) = -\frac{2n}{ka^3}$.

An analysis similar to that in (i) establishes the result.

(iii) $z(a) = ka$, $z'(a) = k$, and $z''(a) = 0$. Equation [2] becomes

$$k_1 = k_2 ,$$

which contradicts the assumption that $k_1 > k_2$. Therefore. the optimal solution is a boundary solution. Suppose the optimal solution has $a_1{}' = 0$ and $a_2{}' > 0$. It will be shown that there is a Pareto superior solution $(a_1{}^*, a_2{}^*)$, with $a_1{}^* > 0$ and $a_2{}^* = 0$. The optimal sharing rule if only task two has nonzero effort is

$$s(x_2) = \left(\lambda + \mu_2 \frac{f_{a_2}(x_2 | a_2{}')}{f(x_2 | a_2{}')} \right)^2 .$$

The Nash condition is (see equation (A2.3) in Appendix 2)

$$2k_2^2 \mu_2 - V'(a_2{}') = 0 .$$

The agent's expected utility is

$$2\lambda - V(a_2{}') = \bar{u} ,$$

and the principal's expected return is (see equation (A2.2) in Appendix 2)

$$k_2 a_2{}' - \lambda^2 - 2\mu_2^2\, k_2^2 .$$

Now consider the pair (a_1^*, a_2^*), where $a_1^* = a_2'$ and $a_2^* = 0$, and consider the sharing rule

$$t(x_1) = \left(\lambda + \mu_1 \frac{f_{a_1}(x_1 | a_1^*)}{f(x_1 | a_1^*)} \right)^2,$$

where $\mu_1 = \dfrac{k_2^2}{k_1^2} \mu_2$. The agent's expected utility (with effort exerted only at task one) is still $\bar{u}$, and the Nash condition is satisfied, since $2k_1^2 \mu_1 = 2k_2^2 \mu_2$ and $a_2' = a_1^*$. Furthermore, the principal is strictly better off because

$$k_1 a_1^* - \lambda^2 - 2\mu_1^2 k_1^2 = k_1 a_2' - \lambda^2 - 2\mu_2^2 k_2^2 \left(\frac{k_2}{k_1} \right)^2$$

$$> k_2 a_2' - \lambda^2 - 2\mu_2^2 k_2^2 .$$

(iv) $z(a) = \ln ka$, $z'(a) = \frac{1}{a}$, and $z''(a) = -\frac{1}{a^2}$. Equations [1] and [2] become

$$\frac{\mu_1^* k_1}{a_1^*} = \frac{\mu_2^* k_2}{a_2^*} \equiv R_p$$

and

$$k_1 + \mu_1^{*2} \cdot \frac{-k_1}{a_1^{*2}} = k_2 + \mu_2^{*2} \cdot \frac{-k_2}{a_2^{*2}},$$

which together imply that

$$k_1 - \frac{R_p^2}{k_1} = k_2 - \frac{R_p^2}{k_2},$$

or

$$R_p^2 \left(\frac{1}{k_1} - \frac{1}{k_2} \right) = k_1 - k_2 > 0$$

since $k_1 > k_2$. Therefore, $\frac{1}{k_1} - \frac{1}{k_2} > 0$, which implies that $k_1 < k_2$ (contradiction). Hence, a boundary solution is optimal.

Suppose the optimal solution has $a_1' = 0$ and $a_2' > 0$. It will be shown that there is a Pareto superior solution (a_1^*, a_2^*), with $a_1^* > 0$ and $a_2^* = 0$. The optimal sharing rule if only task two has nonzero effort is

$$s(x_2) = \left(\lambda + \mu_2 \frac{f_{a_2}(x_2|a_2')}{f(x_2|a_2')} \right)^2 , \quad x_2 = 0,1,\ldots .$$

The Nash condition, evaluated at a_2', is

$$\sum_{x_2=0}^{\infty} \mu_2 \frac{f_{a_2}^2(x_2|a_2')}{f(x_2|a_2')} - V'(a_2') = 0 ,$$

that is, (see Appendix 1),

$$\mu_2(z'(a_2'))^2 \, B''(z(a_2')) - V'(a_2') = 0 ,$$

or $$\mu_2 \cdot \frac{1}{a_2'^2} k_2 a_2' - V'(a_2') = 0 ,$$

which implies that $\dfrac{\mu_2 k_2}{a_2'} - V'(a_2') = 0$.

The agent's expected utility, evaluated at a_2', is

$$2\lambda - V(a_2') = \bar{u} ,$$

and the principal's expected return is

$$k_2 a_2' - \lambda^2 - \frac{\mu_2^2 k_2}{a_2'} .$$

Now consider the pair (a_1^*, a_2^*), where $a_1^* = a_2'$ and $a_2^* = 0$, and consider the following sharing rule,

$$t(x_1) = \lambda + \mu_1 \frac{f_{a_1}(x_1|a_1^*)}{f(x_1|a_1^*)} , \quad x_1 = 0,1,\ldots ,$$

where $\mu_1 = \frac{k_2}{k_1} \mu_2$. The agent's expected utility (with effort exerted only at task one) is still $\bar{u}$, and the Nash condition is satisfied, since $k_1 \mu_1 = k_2 \mu_2$ and $a_1^* = a_2'$. Furthermore, the principal is strictly better off, because

$$k_1 a_1^* - \lambda^2 - \frac{\mu_1^2 k_1}{a_1^*} = k_1 a_2' - \lambda^2 - \frac{k_2^2 \mu_2^2}{k_1 a_2'}$$

$$> k_2 a_2' - \lambda^2 - \frac{\mu_2^2 k_2}{a_2'}.$$

Q.E.D.

<u>Proof of Proposition 3.5.9.</u> The Nash conditions require that

$$\frac{\partial}{\partial a_i} \int 2 \left(\lambda + \sum_{j=1}^{2} \mu_j \left. \frac{\phi_{a_j}(\underline{x}|\underline{a})}{\phi(\underline{x}|\underline{a})} \right|_{\underline{a}^*} \right) \phi(\underline{x}|\underline{a}) d\underline{x} \Bigg|_{\underline{a}^*} - V'(a_1^* + a_2^*) = 0,$$

j=1,2.

For j=1, the condition is

$$2\mu_1 \int \frac{f_{a_1}^2(x_1|a_1^*)}{f(x_1|a_1^*)} dx_1 + 2\mu_2 \int \frac{g_{a_2}(x_2|a_2^*)}{g(x_2|a_2^*)} \cdot f_{a_1}(x_1|a_1^*) g(x_2|a_2^*) d\underline{x}$$

$$- V'(a_1^* + a_2^*) = 0,$$

which reduces to

$$2\mu_1 \int \frac{f_{a_1}^2(x_1|a_1^*)}{f(x_1|a_1^*)} dx_1 = V'(a_1^* + a_2^*) .$$

Since $f(\cdot)$ belongs to Q, the condition can be written as (see Appendix 1)

$$2\mu_1 (z'(a_1^*))^2 B''(z(a_1^*)) = V'(a_1^* + a_2^*) .$$

Since $V' > 0$ and $B''(z(a_1^*)) = Var(x_1|a_1^*) > 0$, $\mu_1^* > 0$. A similar analysis for $j = 2$ shows that $\mu_2^* > 0$.

Appendix 4

First Best

The principal's problem is to

$$\underset{s(\cdot),a_1,a_2(\cdot)}{\text{Maximize}} \quad \int\int W(x-s(x_1,x_2))\phi(x_1,x_2|a_1,a_2(\cdot))\ dx_2dx_1$$

subject to $\int\int [U(s(x_1,x_2))-V(a_1,a_2(\cdot))]\phi(x_1,x_2|a_1,a_2(\cdot))\ dx_2dx_1 \geq \bar{u}$, where $\phi(x_1,x_2|a_1,a_2(\cdot)) = f(x_1|a_1)g(x_2|x_1,a_1,a_2(\cdot))$ and $a_2(\cdot)$ indicates that the agent's second-stage effort is in general not a constant, but rather can depend on any information available at the time of choice. In the scenario described, a_2 may depend on x_1. The Hamiltonian is

$$\int\int W(x-s(\cdot))\phi(\cdot)dx_2dx_1 + \lambda \int\int [U(s(\cdot))-V(\cdot)]\phi(\cdot)dx_2dx_1.$$

Differentiating the Hamiltonian pointwise with respect to s yields

$$-W'\phi + \lambda U'\phi = 0 \text{ for almost every } (x_1,x_2),$$

which implies that

$$\frac{W'(x-s(x_1,x_2))}{U'(s(x_1,x_2))} = \lambda \text{ for almost every } (x_1,x_2). \qquad \text{(A4.1)}$$

1) Risk averse principal, risk neutral agent (i.e., $U' = 1$).

Equation (A4.1) implies that $W'(x-s(x_1,x_2)) = \lambda$ for almost every (x_1,x_2), which implies that $x-s(x_1,x_2)$ is constant for almost every (x_1,x_2), which in turn implies that $s(x_1,x_2) = x-c$, where c is a constant. It will be shown below that a_2 is independent of x_1 if x_1 and x_2 are conditionally independent, in which case $c = E(x|\underline{a}^*) - V(\underline{a}^*) - \bar{u}$.

2) Risk neutral principal, risk averse agent ($W'(x-s(\cdot)) = 1$).

Equation (A4.1) implies that $U'(s(x_1,x_2))$ = constant for almost every (x_1,x_2), which implies that $s(\cdot)$ is a constant for almost every (x_1,x_2). If x_1 and x_2 are conditionally independent, then a_2 is independent of x_1 and $s(\cdot) = U^{-1}(\bar{u} + V(\underline{a}^*))$.

3) Both individuals risk averse.

Equation (A4.1) implies that $x-s(x_1,x_2) = W'^{-1}(\lambda U'(s(x_1,x_2))) \equiv G(s(\underline{x}))$, where $G' > 0$. Therefore, $x = G(s(\underline{x})) + s(\underline{x}) \equiv H(s(\underline{x}))$, where $H' > 0$. Thus, $s(\underline{x}) = H^{-1}(x)$, where $H'^{-1} > 0$.

4) Both individuals risk neutral.

In this case, the agent's expected utility constraint implies that $s(x_1,x_2) = \bar{u} + V(\underline{a}^*)$.

The choice of the agent's effort decisions will first be examined in the simplest case, where the principal is risk neutral, the agent is risk averse, and the outcomes are conditionally independent. That is, $\phi(x_1,x_2|a_1,a_2(\cdot)) = f(x_1|a_1)g(x_2|a_2(\cdot))$, where we allow for the possibility that the second effort decision depends on the first outcome. Since the optimal sharing rule is $s(x_1,x_2) = s$ (constant), the function to be maximized is

$$\int\int (x-s)\,\phi(x_1,x_2|a_1,a_2(\cdot))dx_2dx_1$$
$$+ \lambda[\int\int \{U(s) - V(a_1,a_2(\cdot))\}\phi(x_1,x_2|a_1,a_2(\cdot))dx_2dx_1 - \bar{u}],$$

or, ignoring constants,

$$\int x_1f(x_1|a_1)dx_1 + \int\int x_2f(x_1|a_1)g(x_2|a_2(\cdot))dx_2dx_1$$
$$- \lambda \int\int V(a_1,a_2(\cdot))f(x_1|a_1)g(x_2|a_2(\cdot))dx_2dx_1. \qquad (A4.2)$$

(A4.2) can be rewritten as

$$\int [x_1 + \{\int [x_2 - \lambda\, V(a_1,a_2(\cdot))]g(x_2|a_2(\cdot))dx_2\}]f(x_1|a_1)dx_1. \qquad (A4.3)$$

For each fixed x_1, maximizing the expression inside the braces with respect to a_2 will maximize (A4.3) with respect to a_2. Since the expression depends on x_1 only through a_2, $a_2(\cdot)$ is the same for almost every x_1. That is, a_2

does not depend on x_1. A similar analysis can be done for the case where the principal is risk averse and the agent is risk neutral. Finally, if both individuals are risk averse, then the function to be maximized is

$$\int [\int W(x-s(\underline{x}))g(x_2|a_2(\cdot))dx_2]f(x_1|a_1)dx_1$$

$$+ \lambda [\int \{ \int U(s(\underline{x}))g(x_2|a_2(\cdot))dx_2 - V(a_1,a_2(\cdot))\}f(x_1|a_1)dx_1].$$

In this case, $a_2(\cdot)$ will generally depend on x_1.

Maximizing (A4.3) with respect to a_1 results in the condition that

$$\partial E(x_1|a_1)/\partial a_1 = \lambda \; \partial V(\cdot)/\partial a_1.$$

Maximizing (A4.3) with respect to a_2 (which is independent of x_1) results in the condition that

$$\partial E(x_2|a_2)/\partial a_2 = \lambda \; \partial V(\cdot)/\partial a_2.$$

Proof of Proposition 4.1.1.

Under the given assumptions, $a_2(\cdot)$ will depend on x_1. Let $M_1(a_1)$ denote the mean of x_1 given a_1, and let $M_2(x_1,a_1,a_2(\cdot))$ denote the conditional mean of x_2 with respect to $g(\cdot)$. The function to be maximized is

$$\int \int (x_1 + x_2)\phi(\cdot)dx_2dx_1 - \lambda \int \int V(a_1,a_2(\cdot))\phi(\cdot)dx_2dx_1$$

$$= \int x_1 f(x_1|a_1)dx_1 + \int M_2(\cdot)f(x_1|a_1)dx_1 - \lambda \int V(\cdot)f(x_1|a_1)dx_1$$

$$= M_1(a_1) + E_1M_2(\cdot) - \lambda \, E_1V(\cdot),$$

where E_1 represents expectation with respect to $f(\cdot)$. The first order conditions with respect to effort are then

$$\partial M_1(a_1)/\partial a_1 + \partial E_1M_2(\cdot)/\partial a_1 = \lambda \; \partial E_1V(\cdot)/\partial a_1$$

and

$$\partial M_2(\cdot)/\partial a_2 = \lambda \; \partial V(\cdot)/\partial a_2 \qquad \text{(A4.5)}$$

for almost every x_1 and for $a_1 = a_1^*$. The sign of $a_2^{*\prime}(x_1)$ can be determined by taking the derivative of (A4.5) with respect to x_1. Let the second and third subscripts of j on M_2 denote partial differentiation of M_2 with respect to the j-th argument of $M_2(x_1,a_1,a_2(\cdot))$. Taking the derivative of (A4.5) with respect to x_1 results in

$$M_{233}a_2^{*\prime} + M_{231} = \lambda\ [\partial^2 V(\cdot)/\partial a_2^2]a_2^{*\prime}$$

or

$$a_2^{*\prime}(x_1) = -M_{231}/[M_{233} - \lambda[\partial^2 V(\cdot)/\partial a_2^2]].$$

Q.E.D.

Second Best

Let $\phi(x_1,x_2|a_1,a_2) = f(x_1|a_1)g(x_2|x_1,a_1,a_2(\cdot))$.

The agent's expected utility is

$$\int [\ \int U(s(x_1,x_2))g(x_2|x_1,a_1,a_2(\cdot))dx_2 - V(a_1,a_2(\cdot))]f(x_1|a_1)dx_1.$$

The Hamiltonian is

$$\begin{aligned} H = &\int \int (x-s(x_1,x_2))\phi(\cdot)d\underline{x} \\ &+ \lambda\ \{ \int [\ \int U(s(x_1,x_2))g(x_2|x_1,a_1,a_2(\cdot))dx_2 \\ &\qquad - V(a_1,a_2(\cdot))]f(x_1|a_1)dx_1 - \bar{u}\} \\ &+ \mu_1\{ \int \int U(s(\cdot))[g_{a_1}f + gf_{a_1}]dx_2dx_1 - \int (V_{a_1}f + Vf_{a_1})dx_1\} \\ &+ \int \mu_2(x_1)\ \{ \int U(s(\cdot))g_{a_2}(\cdot)dx_2 - V_{a_2}(\cdot)\}f(x_1|a_1)dx_1. \end{aligned}$$

(a) Differentiating H pointwise with respect to $s(\cdot)$ yields

$$-\phi + \lambda U'\phi + \mu_1 U'\phi_{a_1} + \mu_2(x_1)U'g_{a_2}f = 0 \text{ for almost every } (x_1,x_2).$$

That is,

$$\frac{1}{U'(s(x_1,x_2))} = \lambda + \mu_1 \frac{\phi_{a_1}}{\phi} + \mu_2(x_1)\frac{\phi_{a_2}}{\phi},$$

where the subscript a_2 represents differentiation with respect to a_2 for each fixed x_1.

(b) At $\underline{a} = \underline{a}^*$, $\frac{\partial H}{\partial a_1} = \int\int (x-s(x_1,x_2))\phi_{a_1}(\cdot)d\underline{x}$

$$+ \mu_1 \frac{\partial}{\partial a_1} \{ \int\int\int U(s(\cdot))[g_{a_1} f + gf_{a_1}]dx_2dx_1 - \int (V_{a_1} f + Vf_{a_1})dx_1 \}$$

$$+ \int \mu_2(x_1) \frac{\partial}{\partial a_1} \{ [\int U(s(\cdot))g_{a_2}(\cdot)dx_2 - V_{a_2}(\cdot)]f(x_1|a_1)dx_1 \} = 0.$$

(c) At $\underline{a} = a^*$, and for every fixed x_1,

$$\frac{\partial H}{\partial a_2} = \int (x-s(x_1,x_2))\phi_{a_2}(\cdot)dx_2$$

$$+ \mu_1 \{ \int U(s(\cdot))[g_{a_1a_2} f + g_{a_2} f_{a_1}]dx_2 - (V_{a_1a_2} f + V_{a_2} f_{a_1}) \}$$

$$+ \mu_2(x_1) \{ \int U(s(\cdot))g_{a_2a_2}(\cdot)dx_2 - V_{a_2a_2}(\cdot) \} f(x_1|a_1) = 0.$$

Clearly, the strategy $a_2^*(\cdot)$ depends on x_1 in general. However, if the agent is risk neutral, then the first best solution can be obtained (see Shavell (1979)).

Proof of Proposition 4.2.1: (Generalization of the derivation by Lambert (1981, pp. 104-105).) Since $f(\cdot)$ and $g(\cdot)$ are in Q, they can be written as

$$f(x_1|a) = \exp[z_1(a)x_1 - B_1(z_1(a))]h_1(x_1) \text{ and}$$

$$g(x_2|a) = \exp[z_2(a)x_2 - B_2(z_2(a))]h_2(x_2).$$

Recall that $E(x_i|a) = B_i'(z_i(a))$, $Var(x_i|a) = B_i''(z_i(a))$, and

$f_a/f = z_1'(a)(x_1 - M_1(a_1))$, where $M_1 = E(x_1|a)$. Let

$C_i = \mu_i z_i'(a_i^*)(x_i - M_i(a_i^*))$, where μ_2 denotes $\mu_2(x_1)$. Then $s(\underline{x}) =$

$(\lambda + C_1 + C_2)^2$.

Some helpful quantities will first be calculated.

(1) $\int \int (x_1 + x_2) f(x_1|a_1) g(x_2|a_2(x_1)) dx_1 dx_2$

$$= \int x_1 f(x_1|a_1)dx_1 + \int [\int x_2 g(x_2|a_2(x_1))dx_2] f(x_1|a_1)dx_1$$

$$= M_1(a_1) + \int M_2(a_2(x_1)) f(x_1|a_1)dx_1.$$

(2) $\int \int s(x_1, x_2) f(x_1|a_1) g(x_2|a_2(x_1)) dx_1 dx_2$

$$= \int \int (D + F + G) f(x_1|a_1) g(x_2|a_2(\cdot)) dx_1 dx_2,$$

where $D = (\lambda + C_1)^2$, $F = 2(\lambda + C_1)C_2$, and $G = C_2^2$.

$$E(D) = \int \int [\lambda^2 + 2 \lambda \mu_1 z_1'(a_1^*)(x_1 - M_1(a_1^*))$$
$$+ \mu_1^2 z_1'^2(a_1^*)(x_1 - M_1(a_1^*))^2] f(x_1|a_1) g(x_2|a_2(\cdot)) dx_1 dx_2$$
$$= \lambda^2 + 2 \lambda \mu_1 z_1'(a_1^*)(M_1(a_1) - M_1(a_1^*))$$
$$+ \mu_1^2 z_1'^2(a_1^*)[\mathrm{Var}(x_1|a_1) + M_1^2(a_1) - 2M_1(a_1^*)M_1(a_1) + M_1^2(a_1^*)],$$

since $E(x-a^*)^2 = \mathrm{Var}\, x + (Ex)^2 - 2a^*Ex + a^{*2}$.

$$E(D)\Big|_{\underline{a}^*} = \lambda^2 + \mu_1^2 z_1'^2(a_1^*)\mathrm{Var}(x_1|a_1^*).$$

$$E(F) = 2 \lambda \int \int \mu_2(x_1) z_2'(a_2^*)[x_2 - M_2(a_2^*(\cdot))] f(x_1|a_1) g(x_2|a_2(\cdot)) dx_1 dx_2$$
$$+ 2\mu_1 \int \int z_1'(a_1^*)(x_1 - M_1(a_1^*)) \mu_2(x_1) z_2'(a_2^*(\cdot))(x_2 - M_2(a_2^*(\cdot))) f(\cdot) g(\cdot) dx_1 dx_2$$
$$= 2 \lambda \int \mu_2(x_1) z_2'(a_2^*)[M_2(a_2(x_1)) - M_2(a_2^*(x_1))] f(x_1|a_1) dx_1$$
$$+ 2\mu_1 z_1'(a_1^*) \int z_2'(a_2^*(\cdot)) \mu_2(x_1)(x_1 - M_1(a_1^*))(M_2(a_2(\cdot)) - M_2(a_2^*(\cdot))) f(\cdot) dx_1.$$

$$E(F)\Big|_{\underline{a}^*} = 0.$$

$$E(G) = \int \int \mu_2^2(x_1) z_2'^2(a_2^*(\cdot))(x_2 - M_2(a_2^*(\cdot)))^2 f(\cdot) g(\cdot) dx_1 dx_2$$
$$= \int z_2'^2(a_2^*(\cdot)) \mu_2^2(x_1)[\mathrm{Var}(x_2|a_2(\cdot)) + M_2^2(a_2(\cdot))$$

$$- 2M_2(a_2^*(\cdot))M_2(a_2(\cdot)) + M_2^2(a_2^*(\cdot))]f(\cdot)dx_1.$$

$$E(G)\Big|_{\underline{a}^*} = \int z_2'^2(a_2^*(\cdot))\mu_2^2(x_1)Var(x_2|a_2^*(\cdot))f(x_1|a_1^*)dx_1.$$

Therefore,

$$(2)\Big|_{\underline{a}^*} = \lambda^2 + \mu_1^2 z_1'^2(a_1^*)Var(x_1|a_1^*)$$
$$+ \int z_2'^2(a_2^*(\cdot))\mu_2^2(x_1)Var(x_2|a_2^*(\cdot))f(x_1|a_1^*)dx_1.$$

$$(3)\quad \int\int 2\sqrt{s(\underline{x})}\, f(x_1|a_1)g(x_2|a_2(\cdot))dx_1dx_2 - \int V(a_1,a_2(\cdot))f(x_1|a_1)dx_1$$

$$= 2\int\int [\lambda+\mu_1 z_1'(a_1^*)(x_1-M_1(a_1^*)) + \mu_2(x_1)z_2'(a_2^*(\cdot))(x_2-M_2(a_2^*(\cdot)))]$$
$$\cdot f(\cdot)g(\cdot)dx_1dx_2 - \int V(a_1,a_2(\cdot))f(x_1|a_1)dx_1$$

$$= 2\lambda + 2\mu_1 z_1'(a_1^*)(M_1(a_1) - M_1(a_1^*))$$
$$+ 2\int \mu_2(x_1)z_2'(a_2^*(\cdot))[M_2(a_2(\cdot)) - M_2(a_2^*(\cdot))]f(\cdot)dx_1$$
$$- \int V(a_1,a_2(\cdot))f(x_1|a_1)dx_1.$$

$$(3)\Big|_{\underline{a}^*} = 2\lambda - \int V(a_1,a_2(\cdot))f(x_1|a_1)dx_1 = \bar{u}.$$

$$(4)\quad \frac{\partial(3)}{\partial a_1} = 2\mu_1 z_1'(a_1^*)M_1'(a_1) + 2\int \mu_2(x_1)z_2'(a_2^*(\cdot))[M_2(a_2(\cdot))$$
$$- M_2(a_2^*(\cdot))]f_{a_1}dx_1 - \int V_{a_1}f(x_1|a_1)dx_1 - \int V(\cdot)f_{a_1}(x_1|a_1)dx_1.$$

$$(4)\Big|_{\underline{a}^*} = 2\mu_1 z_1'(a_1^*)M_1'(a_1^*) - \int V_{a_1}f(x_1|a_1^*)dx_1 - \int V(\cdot)f_{a_1}(x_1|a_1^*)dx_1.$$

(5) Fix x_1. $\frac{\partial(3)}{\partial a_2} = 2\mu_2(x_1)z_2'(a_2^*(\cdot))M_2'(a_2(\cdot))f(x_1|a_1)$

$- V_{a_2}(\cdot)f(x_1|a_1) = 0$, which implies that

$$\mu_2(x_1) = \frac{V_{a_2}(\underline{a}^*)}{2z_2'(a_2^*(x_1))M_2'(a_2^*(x_1))}.$$

Clearly, $\mu_2(x_1) > 0$ if $V_{a_2}(\cdot) > 0$. This establishes result (i)(a).

After x_1 is realized, the agent's expected utility given x_1 and $a_2^*(x_1)$ is

$$2 \int \sqrt{s(\underline{x})}\, g(x_2|a_2^*(x_1))dx_2 - V(a_1^*,a_2^*(x_1))$$

$$= 2 \int [\lambda + \mu_1 z_1'(a_1^*)(x_1-M_1(a_1^*)) + \mu_2(x_1)z_2'(a_2^*(x_1))(x_2-M_2(a_2^*(x_1))]$$

$$\cdot\, g(x_2|\cdot)dx_2 - V(a_1^*,a_2^*(x_1))$$

$$= 2[\lambda + \mu_1 z_1'(a_1^*)(x_1-M_1(a_1^*))] - V(a_1^*,a_2^*(x_1)).$$

Differentiating with respect to x_1 yields

$$2\mu_1 z_1'(a_1^*) - V_{a_2}(\cdot)a_2^{*\prime}(x_1).$$

The agent's expected utility for the second stage <u>pecuniary</u> return (i.e., $s(\underline{x})$) is an increasing function of x_1 (assuming $\mu_1 > 0$). Assuming that $V_{a_2}(\cdot) > 0$, a sufficient condition for the agent's expected second stage net utility to be increasing in x_1 is that $a_2^*(x_1)$ be a decreasing function of x_1. This establishes results (i)(b) and (ii)(a).

Now fix x_1 and let a_2 denote $a_2(x_1)$, and f denote $f(x_1|a_1)$.

$$\frac{\partial H}{\partial a_2} = M_2'(a_2)f - 2\lambda\mu_2(x_1)z_2'(a_2^*)M_2'(a_2)f$$

$$- 2\mu_1 z_1'(a_1^*)z_2'(a_2^*)\mu_2(x_1)(x_1-M_1(a_1^*))M_2'(a_2)f$$

$$- z_2'^2(a_2^*)\mu_2^2(x_1)[B_2'''(z_2(a_2))z_2'(a_2) + 2M_2(a_2)M_2'(a_2)$$

$$- 2M_2(a_2^*)M_2'(a_2)]f$$

$$+ \mu_1[2\mu_2(x_1)z_2'(a_2^*)M_2'(a_2)f_{a_1} - V_{a_1a_2}f - V_{a_2}(\cdot)f_{a_1}]$$

$$+ \mu_2(x_1)[2\mu_2(x_1)z_2'(a_2^*)M_2''(a_2)f - V_{a_2a_2}(\cdot)f].$$

$$\frac{1}{f}\cdot\frac{\partial H}{\partial a_2}\bigg|_{\underline{a}^*} = M_2'(a_2^*) - 2\mu_2(x_1)[M_2'(a_2^*)\lambda z_2'(a_2^*) + V_{a_2a_2}/2] - \mu_1 V_{a_2}f_{a_1}/f\bigg|_{\underline{a}^*}$$

$$- \mu_1 V_{a_1a_2}$$

$$+ \mu_2^2(x_1)[2z_2'(a_2^*)M_2''(a_2^*) - z_2'^3(a_2^*)B_2'''(z_2(a_2^*))] = 0.$$

(Note that $f_{a_1}/f\Big|_{\underline{a}^*} - z_1'(a_1^*)(x_1-M_1(a_1^*)) = 0.$)

Substituting the expression for $\mu_2(x_1)$ from (5) above and letting subscripts j on V represent partial differentiation with respect to the j-th effort variable yields

$$M_2' - \lambda V_2 - \frac{V_2V_{22}}{2z_2'M_2'} - \mu_1 V_2 z_1'(x_1-M_1) - \mu_1 V_{12} + \frac{V_2^{\,2}M_2''}{2z_2'M_2'^2} - \frac{V_2^{\,2}z_2'B_2'''}{4M_2'^2} = 0.$$

Differentiating with respect to x_1,

$$M_2''a_2' - \lambda V_{22}a_2' - \frac{a_2'}{2}\left[\frac{V_{22}^{\,2} + V_2V_{222}}{z_2'M_2'} - \frac{V_2V_{22}(z_2''M_2' + z_2'M_2'')}{(z_2'M_2')^2}\right]$$

$$- \mu_1[V_{122}a_2' + V_{22}a_2'z_1'(x_1-M_1) + V_2z_1'] + \frac{a_2'}{2}\left(\frac{2V_2V_{22}M_2'' + V_2^{\,2}M_2'''}{z_2'M_2'^2}\right)$$

$$- \frac{a_2'}{2}\,\frac{V_2^{\,2}M_2''(z_2''M_2'^2 + z_2'2M_2'M_2'')}{(z_2'M_2'^2)^2}$$

$$- \frac{a_2'}{4}\left[\frac{2V_2V_{22}z_2'B_2'''}{M_2'^2} + \frac{V_2^{\,2}z_2''B_2'''}{M_2'^2} + \frac{V_2^{\,2}z_2'^2B_2''''}{M_2'^2} - \frac{V_2^{\,2}z_2'B_2'''2M_2''}{M_2'^3}\right] = 0.$$

Recall that $M_i(a_i) = a_i$, so that $M_i' = 1$ and $M_i'' = 0$. Thus, the expression above reduces to

$$a_2'(x_1) = -\mu_1V_2z_1'/D, \text{ where } D = (\lambda + \mu_1\frac{f_{a_1}}{f})\,V_{22} + \frac{V_{22}^{\,2} + V_2V_{222}}{2z_2'} - \frac{V_2V_{22}z_2''}{2z_2'^2}$$

$$+ \mu_1 V_{122} + \frac{V_2 V_{22} z_2' z_2' B_2''''}{2} + \frac{V_2{}^2 (z_2''B_2''' + z_2'^2 B_2'''')}{4}.$$

Recall that it is assumed that $V_2 > 0$, $V_{22} > 0$, $V_{222} > 0$, and $V_{122} > 0$. It is easily checked that for the exponential, gamma, and Poisson distributions in Q, $z' > 0$, $z'' < 0$, $B'''' > 0$, and $z''B''' + z'^2B'''' \geq 0$. These facts, plus the first order condition requiring that $\lambda + \mu_1 f_{a_1}/f$ be positive, guarantee that the denominator of $a_2'(x_1)$ is positive. The sign of the numerator is the same as the sign of μ_1. Hence, if $\mu_1 > 0$, then $a_2^*(\cdot)$ is a decreasing function of x_1. This establishes result (ii)(b).

Q.E.D.

<u>Proof of Corollary 4.4.1:</u> It is necessary to show that $V_2^* > 0$, $V_{22}^* > 0$, $V_{222}^* > 0$, and $V_{122}^* > 0$. The derivatives of M_i^{-1} will first be calculated. Dropping subscripts for convenience,

1) $M^{-1}(M(a)) = a$ implies that $M^{-1'}(M(a))M'(a) = 1$.

   Therefore, $M^{-1'}(M(a)) = 1/M'(a) > 0$.

2) $M^{-1''}(M(a))M'(a) = - M''(a)/(M'(a))^2$. Therefore,

   $M^{-1''}(M(a)) = - M''(a)/(M'(a))^3 > 0$.

3) $M^{-1'''}(M(a))M'(a) = - [\frac{M'''M'^3 - M''3M'^2M''}{M'^6}]$.

   Therefore, $M^{-1'''}(M(a)) = \frac{3M''^2 - M'''M'}{M'^5}$.

Let subscripts j on V^* denote partial differentiation with respect to e_j. Then

$V_2^* = V'M_2^{-1'} = V'/M_2'(a_2) > 0$,

$$V^*_{22} = V''(M_2^{-1'})^2 + V'M_2^{-1''} = \frac{V''}{(M_2'(a_2))^2} - \frac{V'M_2''(a_2)}{(M_2'(a_2))^3} > 0,$$

$$V^*_{122} = M_1^{-1'}[V'''(M_2^{-1'})^2 + V''M_2^{-1''}]$$

$$= \frac{1}{M_1'(a_1)}\left[\frac{V'''}{(M_2'(a_2))^2} - \frac{V''M_2''(a_2)}{(M_2'(a_2))^3}\right] > 0, \text{ and}$$

$$V^*_{222} = V'''(M_2^{-1'})^3 + V''2M_2^{-1'}M_2^{-1''} + V''M_2^{-1'}M_2^{-1''} + V'M_2^{-1'''}$$

$$= \frac{V'''}{(M'(a_2))^3} - \frac{3V''M_2''(a_2)}{(M_2'(a_2))^4} + \frac{V'(3M_2''^2 - M_2'''M_2')}{M_2'^5}.$$

Thus, if $(3M_2''^2 - M_2'''M_2') > 0$ at a_2^*, then $V^*_{222} > 0$, as required.

Q.E.D.