ANALYSIS OF THE OPTIMAL LINEAR GOAL-BASED INCENTIVE SYSTEM

By HIROYUKI ITAMI*

I. *Introduction*

When a person contracts another person to act (make decision) on his behalf, we say that the agency relationship exists between the two, the former being a principal and the latter an agent of the principal. As a term of the contract, the agent receives some payment (either monetary or non-monetary) in return for his service. This term of the contract that specifies the payment schedule is what is usually called an incentive system in an agency relatoinship. This paper attempts to analyze the optimal linear incentive system in an agency relationship under uncertainty in which the agent has to make two types of decisions, a risk decision and an effort decision, and the incentive system includes a "goal" of the agent's output as well as the output itself as its elements.

The examples of the agency relationships abound in the real world. To name only a few, an employer-employee relationship, a client-attorney relationship, a government-contractor relationship, and a stockholder-manager relationship all have basic elements of an agency relationship. The output-related incentive systems in those relationships are variously called as piece-rate wages, sucess fee schedules, incentive contracts and profit-sharing. Within an organization, the superior's appraisal of the subordinate's performance (and the superior's consequent action like promotion or pay raise) acts as an output-related incentive system.

Basically, an incentive system has two interrelated effects; distributional effect and decisional effect. First of all, any incentive system is a system to divide the output from the agency relationship to concerned parties (the agent(s) and the principal). Different incentive systems mean different distribution, obviously. When the output is uncertain due to environment uncertainty, as is assumed here, different distributions also mean different sharing schemes of risk between the principal and the agent. If the principal and the agent differ in their risk attitudes, it is meaningful to discuss the "optimal" incentive system purely from a distributional viewpoint.

But the matter is more complicated because different distribution further affects the agent's decision (which the agent takes to maximize his (expected) utility) differently. An incentive system has a decisional impact. The optimal incentive system needs balance its distributional effect and decisional effect.

The analysis of the optimal linear incentive system has been attemted, among others,

* Associate Professor (*Jokyōju*) of Management Science.

by Wilson [14], Ross [9], Stiglitz [11]., Stiglitz [12] and Demski and Feltham [4].[1] The linkage of the present paper with these past contributions would become clear as my analysis progresses. The major line of analysis extends Stiglitz's works in that (i) a quite general decision situation (especially the agent output function) is postulated, (ii) the agent decision is divided into a risk decision and effort decision with different characteristics, (iii) a "goal" of the agent's output is incorporated into the incentive system and (iv) both the principal and the agent are risk-averse.

II. *Basic Model*

I assume an agent facing a two-stage sequential decision process under uncertainty. His first-stage decision is x, called risk decision, and the second-stage decision is e, called effort decision. After he has decided on x, he observes s, the state of nature, and he then selects e, given x and and s, with no uncertainty remaining. The agent's output, denoted by z, is a function of x, e and s, $z=f(x, e, s)$. One of the differences between x and e is that x decision is a decision under uncertainty while e is a decision under certainty. Another difference that I assume lies in the effects of these decisions on the agent's utility.

The agent's utility is assumed to depend on the amount of the incentive payment he receives, v, and his effort level, e. Since v generally depends on z, both x and e have effects on the level of v (and thus on the agent's utility). But, the effect of x on the agent's utility is only through v, whereas the effects of e on the utility level is through v and e itself (i.e. disutility of effort).

These differences in the assumptions of two decisions, x and e, are intended to capture some of the essence of what we usually consider as "risk decision" and "effort decision." As a concrete example, consider a division manager of a large corporation as an agent of the headquarter management. He has some decision making authority in investment decision on the projects his division is going to undertake. This is a typical risk decision (x) which he has to make before he knows how the environment turns out. After he has decided, for example, the investment amount on a risky project, he has to implement the project. During this process he is likely to know the environmental condition and exerts effort, if he so wishes, to achieve better output for the organization (z) under the revealed environment. Such effort exertion usually accompanies certain disutility. The effort decision, e, is abstraction of this type of effort exertion. In a general theory of agency relationship, it seems better to assume that the agent as these two types of decisions in a sequential manner.[2]

In order to call x and e as risk decision and effort decision, they have to satisfy

[1] Other related works include Itami [6], Weitzman [13], Mirrlees [7], Aoki [1], Atkins [3] and Harris and Raviv [5].

[2] This does not imply that every agent has both decisions. Some agent may not have any decision authority on a risk decision, although it is fair to assume an effort decision for most agents. Ross [9] assumes a risk decision, but no effort. Stiglitz [11] assumes an elastic effort supply by the agent but his effort is exerted before he knows the environmental condition. More effort increases the variance of the output as well its mean. Is this a very reasonable assumption? Sometimes it may be, but not very often, I think.

certain plausible conditions, especially in their relationships with z. For effort, we perhaps require that more effort brings better output, especially if effort is something that one exerts under certainty with some disutility. Also, decreasing marginal output would be usual. Thus, we may assume

$$(1) \qquad \frac{\partial z}{\partial e} \geqq 0, \qquad \frac{\partial^2 z}{\partial e^2} \leqq 0 \qquad \text{for any } x \text{ and } s.$$

For a risk decision, an essential requirement is that more of a risk decision (x) brings better or worse output depending on the state of nature, s. Furthermore we may require that the more favorable s, the greater the favorable effect of a risk decision. One way to formalize these ideas is to assume the following.

First, we assume

$$(2) \qquad \frac{\partial z}{\partial s} \geqq 0 \qquad \text{for any } x \text{ and } e$$

i.e., s is ordered in such a way that greater s means favorable s. Then, we may formalize the idea of a risk decision by

(3) $\frac{\partial z}{\partial x}$ is an increasing[3] function of s and negative for samll s and positive for large s.

Thus, the marginal effect of greater x is greater if s is more favorable and smaller if s is unfavorable. Then, it is perhaps reasonable to call greater x as a "riskier" decision. Although (1) and (3) are just one of the many possible (and simple) formalization of our intuitive notion of risk decision and effort decision, we will have a chance to use them in the next section to derive a meaningful conclusion on the effects of the parameters of an incentive system.

We write the agent's utility function $A(v, e)$ and assume

$$(4) \qquad A_v > 0, \quad A_e < 0, \quad A_{vv} \leqq 0, \quad A_{ee} \leqq 0, \quad A_{ve} \leqq 0$$

where A_v, A_e, ...are the partial derivatives with respective variables, as usual. As for the principal's utility, we denote the principal's residual, $z-v$, as w and assume that his utility, $P(w)$, depends only on w. Furthermore, we assume

$$(5) \qquad P_w > 0 \text{ and } P_{ww} \leqq 0$$

where P_w and P_{ww} are the first and second derivatives of P.

The principal's decision is concerned with the determination of the incentive function v. The incentive payment can be related not only to the agent's output (z), but also to other variables indicative of the agent's decision, ability and so forth. In this paper, I consider a liner, goal-based incentive system, i.e.

$$(6) \qquad v = a + bz + cz_0$$

where a, b, c are incentive parameters and z_0 is the output goal for the agent. The way this goal level is set can vary depending on the situation. The agent may set z_0 for himself and then report it to the principal, possibly with some bias. Or, the principal

[3] Whenever I use "increasing" or "greater" etc., its precise meaning is "non-decreasing" or "greater or equal to," etc. To avoid awkward wording I use a simpler word unless it is confusing.

may give a given level of output as the agent's goal for attainment. Since I do not treat the goal setting process endogenously within the framework of this paper, I simply assume that z_0 is a function of x (and incentive parameters), including a special case where z_0 is a given constant.[4] When z_0 varies with x and incentive parameters, this perhaps corresponds to the case where the agent sets his own goal.

It is necessary to note that the goal setting (and reporting) process is an important consideration in incentive system design, especially if the agent plays a key role in setting the goal. One of the objectives of the design of an incentive system is often to influence this goal-setting process so that the principal can gather better information regarding the agent's decision environment, attitude, ability and so on through the self-set goal.[5]

The principal may desire to include the goal level into the incentive system as a motivational device to influence the agent's decision, especially an effort decision. A special form of (6) is often used with this objective in mind, i.e.,

$$(7) \qquad v = a + b'z + c'z_0 + r(z - z_0)$$

where r is a constant reward parameter. In this scheme, the agent's incentive is determind partly by how much he overshoot (or undershoot) the set goal. The principal's hope is that the agent exerts more effort to avoid penalty for goal underachievement ($z-z_0<0$) and to be rewarded for goal overachievement ($z-z_0<0$). This motivational impact would be greater when $z-z_0$ determines v in a nonlinear fashion (for example, the penalty rate for underachievement is greater than the reward rate for overachievement). Although the nonlinear case seems more realistic, I will analyze a nonliner, goal-based incentive system in a future paper. Because of the linearity, it is evident that (7) can be transformed to (6) by defining $b=b'+r$ and $c=c'-r$.

Given this incentive system, the agent has to decide x and then e. The optimality condition for the effort decision is, for a given x and s,

$$(8) \qquad A_v \cdot \frac{\partial v}{\partial e} + A_e = 0$$

The risk decision is a little more complicated. When the agent maximizes his expected utility, he has to consider not only the direct effect of x on v (through z and z_0) but also the effect of x on the effort decision e (and then to v). x becomes a parameter in the effort decision process. Denote by $\frac{\partial e}{\partial x}$ the partial derivative of the optimal effort function $e=e^\circ(x, s)$ which is implicitly given by (8), assuming differentiability. Then, the expected utility maximization leads to

$$E\left(A_v\left(\frac{\partial v}{\partial x} + \frac{\partial v}{\partial e} \cdot \frac{\partial e}{\partial x}\right) + A_e \cdot \frac{\partial e}{\partial x}\right) = 0$$

where E is the expectation with respect to s. Using (8), we obtain

$$(9) \qquad EA_v \cdot \frac{\partial v}{\partial x} = 0$$

Note that in (9) e is the optimal effort $e^\circ(x, s)$. It is likely that x has to satisfy a certain

[4] Note that z_0 does not depend either on e or s.

[5] For some discussion on this point, see Weitzman [13] and Demski and Feltham [4].

constraint which we do not treat explicitly above. For example, if x is the amount to be invested in a risky project, x has to be within the available fund. In the following, we assume the optimal x occurs in the interior of the contraint set.

III. *Risk Decision Effect and Effort Decision Effect*

In discussing the effect of a linear incentive system, especially the effect of b (sharing parameter), it is often said that b has two types of effects, riks-sharing and incentive effects. A large b forces the agent to share too much risk from the uncertainty of z, but at the same time has motivational effects which may lead to more effort. A small b affects the agent in just the opposite way.[6] If the incentive system forces the agent to share a large risk, it may be undesirable for the agent from the purely distributional point of view. Furthermore, it may cause the agent to take too conservative a decision and thus undesirable from the principal's viewpoint as well, even if it has certain motivational effect in terms of effort.

An implicit supposition in this line of argument is clear. Greater b will make the agent decision more conservative but induce greater effort exertion. But is this really true under reasonable assumptions? The purpose of this section is to discuss the decisional effects of incentive parameters of (6) within the framework of our basic model. The insights we could obtain from this analysis will aid us in understanding how the optimal balance will be struck among the various effects of the incentive parameters in determining the incentive system, a topic of the next section.

We first derive the basic comparative static equations for our sequential decision process. Let y be an incentive parameter ($y=a$, b, or c). From (8) and (9), we have

$$G(x, y) = 0$$
$$F(x, e, y, s) = 0 \qquad \text{for each } s$$

where $G(x, y)=E\dfrac{\partial A}{\partial x}$ (e is $e^{\circ}(x, y, s)$ to be determined from (9), and $F(x, e, y, s)=\dfrac{\partial A}{\partial e}$. When y changes infinitesimally, both x and e change satisfying (8)and (9) *simultaneously*. Thus, we have the simultaneous equations for the total differentials, dy, dx,and de:

(10) $$G_x \cdot dx + G_y \cdot dy = 0$$

(11) $$F_x \cdot dx + F_e \cdot de + F_y \cdot dy = 0 \qquad \text{for each } s$$

From this we obtain

$$\frac{dx}{dy} = -\frac{G_y}{G_x}$$
$$\frac{de}{dy} = -\frac{F_y}{F_e} + \frac{dx}{dy}\left(-\frac{F_x}{F_e}\right)$$

where $\dfrac{dx}{dy}$, $\dfrac{de}{dy}$ are respectively the rate of changes of the decisions in response to

[6] See, for example, Stiglitz [11].

changes in y, incorporating the simultaneous effects of y on x and e in the two-stage decision process.[7]

Let us denote by H_{xx} a usual second partial derivative of $A(v, e)$ with respect to x without considering x's effect on e and define H_{xe}, H_{ey}, etc. similarly. Let us also define $\frac{\partial e}{\partial y}$, $\frac{\partial e}{\partial x}$ as the changes in e in response to changes in y and x in the second-stage decision disregarding the interaction with the first-stage decision. Let $\frac{\partial x}{\partial y}$ be the changes in x in response to a change in y in the first-stage decision, disregarding the kick-back effect from $\frac{\partial e}{\partial y}$ in the second-stage decision. Then,

$$G_x = E\left(H_{xx}+H_{ex}\cdot\frac{\partial e}{\partial x}\right)$$

$$G_y = E\left(H_{xy}+H_{xe}\cdot\frac{\partial e}{\partial y}\right)$$

$$\frac{\partial e}{\partial y} = -\frac{H_{ey}}{H_{ee}} = -\frac{F_y}{F_e}$$

$$\frac{\partial e}{\partial x} = -\frac{H_{ex}}{H_{ee}} = -\frac{F_x}{F_e}$$

$$\frac{\partial x}{\partial y} = -\frac{E(H_{xy})}{E\left(H_{xx}-\frac{(H_{ex})^2}{H_{ee}}\right)}$$

Thus, we finally obtain

$$\frac{dx}{dy} = \frac{\partial x}{\partial y} + \frac{E\left(H_{xe}\cdot\frac{\partial e}{\partial y}\right)}{E\left(H_{xx}-\frac{(H_{ex})^2}{H_{ee}}\right)} \tag{12}$$

$$\frac{de}{dy} = \frac{\partial e}{\partial y} + \frac{dx}{dy}\cdot\frac{\partial e}{\partial x} \tag{13}$$

These equations are, in a sense, like the Slutsky equations in consumer's demand theory. The first term on the RHS of (12) and (13) indicates the direct effects of a parametric change of y and the second-term measures the repurcussion of this parametric change through sequential decisions. As with the Slutsky equations, we will see that the sign of the first term is determinate for x and often determinate with some qualifications for e. The second term, however, can be positive or negative. There is no simple way

[7] Here $\frac{dx}{dy}$, $\frac{de}{dy}$ are used to mean the effect of y on x or e fully considering the kickback effects of changes in y through a two-stage decision process, to distinguish them from $\frac{\partial x}{\partial y}$ or $\frac{\partial e}{\partial y}$ below.

of knowing. The total effects, $\frac{dx}{dy}$ and $\frac{de}{dy}$, are thus positive or negative for $y=a$, b or c. The answer to our initail question (e.g., Does greater b lead to more conservative risk decision and greater effort exertion?) is not clear in any definitive way.

But, we can derive much more definitive answers at least about the direct effect of incentive parameters, $\frac{\partial x}{\partial y}$ and $\frac{\partial e}{\partial y}$. This is our main task here. To do so, we assume the following in this section, on top of the assumptions ((1)~(4)) made in the previous section.

(14) $b \geqq 0, \quad c \geqq 0, \quad a + cz^\circ \geqq 0$

(15) $R_A = -\frac{A_{vv}}{A_v}$ does not depend on e

(16) R_A is decreasing in v, and vR_A is increasing in v.

(17) $\frac{\partial z_\circ}{\partial x} \geqq 0$, $z_\circ \geqq 0$, $z_\circ$ does not depend on a, b or c.

(18) $\left.\frac{\partial v}{\partial x}\right|_{x=x^\circ}$, as a function of s through $z=f(x^\circ, e^\circ(x^\circ, s), s)$, changes sign from negative to positive only once as s increases.

x° and e° denote the optimal decisions.

(19) $\frac{\partial^2 z}{\partial e \partial s} \geqq 0$

The assumption (16) is a familiar assumption of decreasing absolute risk aversion and increasing relative risk aversion. To assume this in our framework where v is related to e and e is an argument in both A_v and A_{vv}, the assumption (15) is necessary.[8]

The assumption (18) perhaps needs some explanation. Evaluated at $x=x^\circ$, $\frac{\partial v}{\partial x}$ nonetheless varies as s changes through two routes. First, s has direct effect on $\frac{\partial z}{\partial x}$ because s is the third argument in f function. Second, as s varies, the agent responds by changing his effort level and thus $\frac{\partial v}{\partial x}$ changes further. The assumption (18) says that the net result of these two effects of s on $\frac{\partial v}{\partial x}$ is that $\frac{\partial v}{\partial x}$ changes its sign only once. Thus, if $\frac{\partial v}{\partial x} > 0$ for some s, then $\frac{\partial v}{\partial x} > 0$ for all s which is greater than this s. A sufficient condition for the assumption (18) is that $\frac{\partial v}{\partial x}$ is an increasing function of s. Since, $\frac{\partial^2 z}{\partial x \partial s} \geqq 0$ and

[8] This is satisfied if the utility function $A(v, e)$ is of the form
$$A(v, e) = C_\circ + \alpha_1 C(v) + \alpha_2 C_2(e) + \alpha_3 C_1(v) C_2(e).$$

$$\frac{\partial^2 v}{\partial x \partial s} = b\left(\frac{\partial^2 z}{\partial x \partial s} + \frac{\partial^2 z}{\partial x \partial e} \cdot \frac{\partial e}{\partial s}\right)$$

where $\frac{\partial e}{\partial s}$ is the rate of changes of e in response to s in the second-stage decision, the monotonicity of $\frac{\partial v}{\partial x}$ is guaranteed if $\frac{\partial^2 z}{\partial x \partial e} \cdot \frac{\partial e}{\partial s}$ is either non-negative or of sufficiently small magnitude when it is negative. An example is $\frac{\partial^2 z}{\partial x \partial e} = 0$.

The assumption (19) is reasonable, meaning under a more favorable state, the marginal output of effort is greater.

Given these assumptions, we now prove a lemma concerning the level of incentive payment in relation to s. Let $v(s)$ be

$$v(s) = a + bf(x^\circ, e(s), s) + cz^\circ \tag{20}$$

where $e(s)$ is the optimal effort given x° and s. This is the incentive payment the agent receives with x° and s.

LEMMA 1. $v(s)$ is increeasing with respect to s.

Proof. Suppose $s_1 \geqq s_2$ and let e^* such that

$$f(x^\circ, e^*, s_2) = f(x^\circ, e(s_1,) s_1) \tag{21}$$

Since $\frac{\partial z}{\partial e} \geqq 0$, $\frac{\partial z}{\partial s} \geqq 0$, we have

$$e^* \geqq e(s_1) \tag{22}$$

Let $(\cdot)_{s=s_2}^{e=e^*}$ denote the value of a function in the parenthesis evaluated at the indicated values of e and s.

We note that A_v and A_e are functions of v and e and do not depend explicitly on s. Thus, using (21), (22), $A_{ve} \leqq 0$ and $A_{ee} \leqq 0$, we have

$$(A_v)_{s=s_2}^{e=e^*} \leqq (A_v)_{s=s_1}^{e=e(s_1)}$$

$$(A_e)_{s=s_2}^{e=e^*} \leqq (A_e)_{s=s_1}^{e=e(s_1)}$$

By using $\frac{\partial^2 z}{\partial e^2} \leqq 0$ and $\frac{\partial^2 z}{\partial e \partial s} \geqq 0$, we obtain

$$\left(\frac{\partial v}{\partial e}\right)_{s=s_2}^{e=e^*} \leqq \left(\frac{\partial v}{\partial e}\right)_{s=s_1}^{e=e^*} \leqq \left(\frac{\partial v}{\partial e}\right)_{s=s_1}^{e=e(s_1)}$$

Combining these inequalities,

$$\left(A_v \cdot \frac{\partial v}{\partial e} + A_e\right)_{s=s_2}^{e=e^*} \leqq \left(A_v \cdot \frac{\partial v}{\partial e} + A_e\right)_{s=s_1}^{e=e(s_1)} = 0$$

Then, it follows, from the concavity of $A(v, e)$ with respect to e,

$$e^* \geqq e(s_2)$$

From this and (21), we have

$$v(s_2)\leqq v(s_1)$$

Thus, $v(s)$ is increasing with respect to s.

PROPOSITION 1. (Direct Risk Decision Effect)

$$\frac{\partial x}{\partial a}\geqq 0 \tag{23}$$

$$\frac{\partial x}{\partial b}\leqq 0 \tag{24}$$

$$\frac{\partial x}{\partial c}\geqq 0 \tag{25}$$

Proof. From the second-order condition of the optimality of the first-stage decision, we have

$$E\left(H_{xx}-\frac{(H_{xe})^2}{H_{ee}}\right)<0$$

Then,

$$\frac{\partial x}{\partial y}\sim E(H_{xy}) \tag{26}$$

where "$\sim$" means "of the same sign as." And,

$$H_{xa}=-A_v\cdot\frac{\partial v}{\partial x}\cdot R_A \tag{27}$$

$$H_{xb}=-A_v\cdot\frac{\partial v}{\partial x}\cdot zR_A+A_v\cdot\frac{\partial z}{\partial x}=zH_{xa}+A_v\cdot\frac{\partial z}{\partial x} \tag{28}$$

$$H_{xc}=-A_v\cdot\frac{\partial v}{\partial x}\cdot z_\circ R_A+A_v\cdot\frac{\partial z_\circ}{\partial x}=z_\circ H_{xa}+A_v\cdot\frac{\partial z_\circ}{\partial x} \tag{29}$$

From (18) there exists $s_\circ$ such that

$$\frac{\partial v}{\partial x}\geqq 0 \qquad \text{for } s\geqq s_\circ$$

$$\frac{\partial v}{\partial x}\leqq 0 \qquad \text{for } s\leqq s_\circ$$

Since $v(s)$ is increasing in s, R_A is decreasing in s. Let R_A° be the value of R_A at $s=s_\circ$ (i.e., $v=v(s_\circ)$). Then, we have

$$A_v\cdot\frac{\partial v}{\partial x}\cdot R_A\leqq A_v\cdot\frac{\partial v}{\partial x}R_A^\circ \qquad \text{for all } s \tag{30}$$

Taking the expected value of (30) and using $EA_v\cdot\frac{\partial v}{\partial x}=0$ at the optimal x, we obtain

$$E\left(A_v \cdot \frac{\partial v}{\partial x} \cdot R_A\right) \leqq 0 \tag{31}$$

This gives us (23).

If $b \neq 0, z = \frac{1}{b}v - \frac{1}{b}(a + cz_0)$. Then

$$zR_A = \frac{1}{b}(vR_A - (a + cz_0)R_A)$$

Since vR_A is increasing and R_A decreasing in v, it follows

(32) zR_A is increasing in s (and z)

Using similar argument as (31), we get

$$E\left(A_v \cdot \frac{\partial v}{\partial x} \cdot z \cdot R_A\right) \geqq 0$$

From $EA_v \cdot \frac{\partial v}{\partial x} = 0$,

$$bEA_v \cdot \frac{\partial z}{\partial x} = -cEA_v \cdot \frac{\partial z_0}{\partial x} \leqq 0$$

Thus, we have (24). The case $b=0$ is trivial.

Since $z_0 \geqq 0$, $\frac{\partial z_0}{\partial x} \geqq 0$, (29) gives us (25) by noting (31).

From the way x is assumed to affect z, it is reasonable to say that greater x means more risk-taking. The results in this proposition closely parallel those obtained in the comparative static analysis of risk-taking effects in portfolio analysis and in the theory of income taxation.[9] Increasing a is similar to increasing initial wealth and increasing b is similar to decreasing linear income tax rate. Increasing wealth is usually associated with more risk-taking and decreasing tax rate is associated with less risk-taking, as is implied in this proposition. Our results indicate that these earlier results are quite general, although we should be careful to note that these results are concerned only with the direct risk-taking effect, not the total effect.

The fact that c has a risk-encouraging effect $\left(\frac{\partial x}{\partial c} \geqq 0\right)$ is quite reasonale from the assumptions. Since $\frac{\partial z_0}{\partial x} \geqq 0$, greater weight on z_0 would lead to greater x. Greater c also implies an increase in non-risky incentive payment $(a+cz_0)$ and thus has the same effect as in the case of a. Becasuse of these dual effects, c plays a pivotal role in optimal risk encouragement as we shall see in the next section.

Although the risk decision effect is quite clear-cut for all the incentive parameters, the effort decision effect is less definitive. The effort effect of b can be positive or negative depending on the situation. This is reminiscent of the backward bending supply curve of labor. Increasing a sharing parameter (or wage in labor theory) lead to more effort (or labor supply) up to a point, but beyond some point greater b leads to less

[9] See, for example, Arrow [2] (Chapter 3) and Stiglitz [10].

effort. These are the results we prove in:

PROPOSITION 2. (Effort Decision Effect)

$$\frac{\partial e}{\partial a} \leqq 0 \tag{33}$$

$$\frac{\partial e}{\partial b} \geqq 0 \qquad \text{if } bz \leqq \frac{1}{(1+t)R_A} \tag{34}$$

$$\frac{\partial e}{\partial b} \leqq 0 \qquad \text{if } bz \geqq \frac{1}{(1+t)R_A}$$

where

$$t = \frac{A_{ve}}{A_{vv} \cdot \dfrac{\partial v}{\partial e}} \geqq 0$$

$$\frac{\partial e}{\partial c} \leqq 0 \tag{35}$$

Proof. Since $H_{ee}<0$ frm the concavity of $A(v, e)$ in e,

$$\frac{\partial e}{\partial y} \sim H_{ey}$$

We then have, after suitable rearrangement,

$$H_{ea} = -A_v \cdot \frac{\partial v}{\partial e}(1+t)R_A \tag{36}$$

$$H_{eb} = A_v \cdot \frac{\partial z}{\partial e}\{1 - b(1+t)zR_A\} = zH_{ea} + A_v \cdot \frac{\partial z}{\partial e} \tag{37}$$

$$H_{ee} = -A_v \cdot \frac{\partial z}{\partial e} b(1+t)z_0 R_A = z_0 H_{ea} \tag{38}$$

Then, (33), (34) and (35) follows immediately.

These results are intuitively reasonable. When the performance-independent reward $(a+cz_0)$ is increased, the agent has less motivation to exert effort. Increase in b induces the agent to work harder unless b is too gerat and he feel "saturated" with the income in comparison with the disutility of more effort. The conditions in (34) can be converted, if $z>0$, to

$$\frac{\partial e}{\partial b} \geqq 0 \qquad \text{if } b \leqq \frac{1}{(1+t)zR_A}$$

$$\frac{\partial e}{\partial b} \leqq 0 \qquad \text{if } b \geqq \frac{1}{(1+t)zR_A}.$$

Since zR_A is increasing in s as in (32), the above conditions implies that, unless t changes substantially as s varies, the less favorable the state of nature (smaller s), the more likely that greater b induces more effort. This is a reasonable conclusion. It is also interesting to note that

$$\frac{1}{1+t}=\frac{A_{vv}\cdot\frac{\partial v}{\partial e}}{A_{vv}\cdot\frac{\partial v}{\partial e}+A_{ve}}.$$

Thus, $\frac{1}{1+t}$ is the proportion of decrease in marginal utility of income (by increasing effort) due to "income saturation." If this proportion is greater, the more likely $\frac{\partial e}{\partial b}>0$.

In the previous section, I mentioned a well-practiced incentive system which depends on the deviation of the actual performance from the set goal, (7). In light of the results in Proposition 1 and Proposition 2, we can have better insights into the direct decision effects of changing a deviation incentive parameter (r). An increase of r means a simultaneous increase of b and decrease of c in Propositions 1 and 2 and thus

$$\frac{\partial x}{\partial r}=\frac{\partial x}{\partial b}-\frac{\partial x}{\partial c}$$

$$\frac{\partial e}{\partial r}=\frac{\partial e}{\partial b}-\frac{\partial e}{\partial c}$$

We can then see

$$\frac{\partial x}{\partial r}\leqq\frac{\partial x}{\partial b}\leqq 0 \tag{39}$$

$$\frac{\partial e}{\partial r}=-A_v\cdot\frac{\partial z}{\partial e}\{1-b(1+t)(z-z_0)R_A\}\geqq\frac{\partial e}{\partial b} \tag{40}$$

(39) implies that greater r causes more concervative risk decision, but is likely to have an effort motivating effect. At least the effort motivating effect of r is greater than that of b and $\frac{\partial e}{\partial r}$ is positive as long as $z<z_0$, i.e., when the environment is unfavorable so that the goal is not achievable with the present level of effort. It is sometimes argued that a gerat emphasis on the deviation $(z-z_0)$ may motivate more effort, but at the same time often implants too much conservatism in the agent's risk decision.[10] The results in (39) and (40) certainly underwrite this line of reasoning, as far as the direct decision effects are concerned.

Example. To see the results obtained so far a little more concretely and also to indicate the kinds of the relationships between the total decision effects and the direct decision effects, let us analyze the following simple example. Suppose

$$z=sx+e \quad \text{with} \quad E(s)>0,\ 1\geqq x\geqq 0$$

$$v=a+bz \quad \left(\text{or } \frac{\partial z_0}{\partial x}=0\right)$$

$$A_{ve}=0, \quad R=-\frac{A_{ve}}{A_v} \text{ and } Q=\frac{A_{ee}}{A_e} \text{ are constant.}$$

[10] See Itami [6] for an analysis of risk-taking by the agent under a non-linear deviation-based incentive system.

The optimality conditions for the first-stage and the second-stage are, respectively,

(41) $EA_v \cdot s = 0$

(42) $A_v \cdot b + A_e = 0 \quad \text{or} \quad b = -\frac{A_e}{A_v}.$

The second derivatives of $A(v, e)$ are

$$H_{xx} = A_{vv}(bs)^2,\ H_{xe} = A_{vv}b^2s,\ H_{ee} = A_{vv}b^2 + A_{ee}$$
$$H_{xa} = -b \cdot A_v \cdot sR,\ H_{xb} = A_v \cdot s(1 - bzR)$$
$$H_{ea} = A_{vv} \cdot b,\ H_{eb} = A_v(1 - bzR).$$

We also have, by using (42) at the optimum,

$$H_{xa} - H_{xe} \cdot \frac{H_{ea}}{H_{ee}} = -\frac{bQR}{Q + bR} A_v s$$

$$H_{xb} - H_{xe} \cdot \frac{H_{eb}}{H_{ee}} = \frac{Q}{Q + bR}(1 - bzR)A_v s$$

$$\frac{H_{xe}}{H_{ee}} = \frac{bR}{Q + bR} s$$

$$\frac{H_{ea}}{H_{ee}} = \frac{R}{Q + bR}$$

$$\frac{H_{eb}}{H_{ee}} = \frac{1}{(Q + bR)b}(1 - bzR).$$

Therefore, using (41) and $E(s) > 0$

$$\frac{dx}{da} = \frac{\partial x}{\partial a} = 0$$

(43) $$0 \geqq \frac{dx}{db} = \frac{Q}{Q + bR} \frac{\partial x}{\partial b} \geqq \frac{\partial x}{\partial b}$$

(44) $$E\left(\frac{\partial e}{\partial x}\right) < 0.$$

Furthermore,

$$E\left(\frac{de}{da}\right) = E\left(\frac{\partial e}{\partial a}\right) < 0$$

(45) $$E\left(\frac{de}{db}\right) \geqq E\left(\frac{\partial e}{\partial b}\right).$$

Especially interesting are (43), (44) and (45). (44) means that x and e are, on the average, substitutes. More risky decision tend to accomany less effort. (43) and (45) give us some idea about the difference between the direct decision effects and the total effects of b. For both the risk decision effect and the effort decision effect, the total effect is greater than the direct effect. Thus, in a conventional analysis in which simultaneous consideration of risk decision and effort decision is lacking, the magnitude of the

negative risk effect is overestimated by ignoring the effort decision. Since $E\left(\frac{\partial e}{\partial b}\right)>0$ is likely, the magnitude of the positive effort effect is underestimated by ignoring the risk decision. The upshot is that the optimal b would be greater if there are both risk and effort decision than the case where only either decision is explicitly considered in incentive system design. Needless to say, this observation is based on the analysis of a simple example and is only, at most, suggestive of a general tendency.

IV. *Optimal Incentive System*

Let us suppose that the agent is a utility-taker and that the principal selects the optimal incentive system which maximizes his own expected utility, subject to the constraint that the agent is guaranteed (at least) a given level of the expected utility when the agent has optimally chosen x and e. More formally, the optimal incentive system is the solution of:

$$
\begin{aligned}
(46) \qquad & \max_{a,b,c} EP(w) \\
& \text{s.t.} \quad EA(v)\geqq\bar{A} \\
& \qquad v=a+b\ f(x^{\circ},e^{\circ},s)+cz_{\circ}
\end{aligned}
$$

where x° and e° are the agent's optimal decisions defined in the previous section.

This is one of the possible concepts of "optimality" of an incentive system in the agency relationship. It treats the agency relationship essentially as a non-cooperative game and reduces to a special case of a Nash-equilibrium. Ross [8], Stiglitz [11], Mirrlees [7] use this concept of optimality. Another concept of optimality that has appeared in the literature is that of Pareto optimality. Pareto optimality in this case is defined as Pareto optimality in a cooperative game. An incentive system is judged to be optimal if it enables, in a sense, to transform the agency relationship to a cooperative game, even though the x and e decision are still chosen by the agent to maximize his own utility. Wilson [14] and Ross [9] have discussions based on this optimality concept. Pareto optimality is very nice, but often hard to obtain. Since it seems natural to consider that the basic character of the agency relationship contains an element of a non-cooperative game, I take the principal's maximization above as optimality in my framework. An investigation of the relationship among the various concepts of optimality is a topic of further research.[11]

In deriving the optimal incentive system and discussing its property, we assume the assumptions (1), (2), (4), (5), (17), (19). The assumption (17) is not essential, but simplifies our discussion. It is an easy exercise to see that $z_{\circ}$ can depend on b and c in the following without altering the conclusions. In the assumption (4) we will sometimes require $A_{ve}=0$ and explicitly note this when necessary.

Because of the utility level constraint in (46), the principal cannot select a, b, c

[11] Ross [9] discusses the relationships between the two principles (similarity and Pareto-efficiency) of incentive systems in an agency.

independently. Since the inequality constraint reduces to the equality constraint, this constraint implicitly gives a as a function of b and c. Following Stiglitz [11], let us call a triplet (a, b, c) that satisfies this constraint as a utility-equivalent incentive system. Among the utility-equivalent incentive system, we have

$$\frac{\partial a}{\partial b} = -\frac{EA_v z}{EA_v}, \quad \frac{\partial a}{\partial c} = -z_0 \tag{47}$$

Selecting among the utility-equivalent incentive systems, the principal's optimization in (46) gives:

$$EP_w\left\{\left(\frac{\partial w}{\partial b}\right)_{\bar{A}} + \frac{\partial w}{\partial x}\left(\frac{dx}{db}\right)_{\bar{A}} + \frac{\partial w}{\partial e}\left(\frac{de}{db}\right)_{\bar{A}}\right\} = 0 \tag{48}$$

$$EP_w\left\{\left(\frac{\partial w}{\partial c}\right)_{\bar{A}} + \frac{\partial w}{\partial x}\left(\frac{dx}{dc}\right)_{\bar{A}} + \frac{\partial w}{\partial e}\left(\frac{de}{dc}\right)_{\bar{A}}\right\} = 0 \tag{49}$$

where $\left(\frac{\partial w}{\partial b}\right)_{\bar{A}}$, $\left(\frac{dx}{db}\right)_{\bar{A}}$ ··· are utility-equivalent derivatives. For example,

$$\left(\frac{dx}{db}\right)_{\bar{A}} = \frac{dx}{db} + \frac{dx}{da}\cdot\frac{\partial a}{\partial b}.$$

In particular, we observe for $y=b$ and c, using (13):

$$\left(\frac{de}{dy}\right)_{\bar{A}} = \frac{de}{dy} + \frac{de}{da}\cdot\frac{\partial a}{\partial y} = \left(\frac{\partial e}{\partial y}\right)_{\bar{A}} + \left(\frac{dx}{dy}\right)_{\bar{A}}\frac{\partial e}{\partial x} \tag{50}$$

where

$$\left(\frac{\partial e}{\partial y}\right)_{\bar{A}} = \frac{\partial e}{\partial y} + \frac{\partial e}{\partial a}\cdot\frac{\partial a}{\partial y}.$$

Using (47) and (50) we obtain,

$$\left(\frac{dx}{dc}\right)_{\bar{A}} = -\frac{1}{E\left(H_{xx} - \frac{(H_{ex})^2}{H_{ee}}\right)}\cdot EA_v\cdot\frac{\partial z_0}{\partial x} \tag{51}$$

$$\left(\frac{de}{dc}\right)_{\bar{A}} = \left(\frac{dx}{dc}\right)_{\bar{A}}\cdot\frac{\partial e}{\partial x} \tag{52}$$

From (47), (50), (51), (52) we can rearrange (48), (49):

$$EP_w\left[\left\{(1-b)\left(\frac{\partial z}{\partial x} + \frac{\partial z}{\partial e}\cdot\frac{\partial e}{\partial x}\right) - c\frac{\partial z_0}{\partial x}\right\}\left(\frac{dx}{db}\right)_{\bar{A}} + (1-b)\frac{\partial z}{\partial e}\left(\frac{\partial e}{\partial b}\right)_{\bar{A}}\right] = EP_w\left(z - \frac{EA_v z}{EA_v}\right) \tag{53}$$

$$EP_w\left\{(1-b)\left(\frac{\partial z}{\partial x} + \frac{\partial z}{\partial e}\cdot\frac{\partial e}{\partial x}\right) - c\frac{\partial z_0}{\partial x}\right\}\left(\frac{dx}{dc}\right)_{\bar{A}} = 0 \tag{54}$$

Thus, (51) and (54) imply:

$$(55)\qquad \begin{cases} EP_w\left\{(1+b)\left(\dfrac{\partial z}{\partial x}+\dfrac{\partial z}{\partial e}\cdot\dfrac{\partial e}{\partial x}\right)-c\dfrac{\partial z_\circ}{\partial x}\right\}=\dfrac{\partial}{\partial x}EP=0 \quad \text{or} \\ \dfrac{\partial z_\circ}{\partial x}=0. \end{cases}$$

Let, assuming $E(z)\neq 0$:

$$(56)\qquad q_P=\frac{P_w}{EP_w},\quad q_A=\frac{A_v}{EA_v}$$

$$(57)\qquad \theta_P=1-\frac{EP_w\cdot z}{E(z)EP_w},\quad \theta_A=1-\frac{EA_v\cdot z}{E(z)EA_v}$$

$$(58)\qquad \begin{cases} m=0 & \text{if } \dfrac{\partial z_\circ}{\partial x}>0 \\ m=Eq_P\cdot\left(\dfrac{\partial z}{\partial x}+\dfrac{\partial z}{\partial e}\cdot\dfrac{\partial e}{\partial x}\right)\left(\dfrac{dx}{db}\right)_{\bar{A}} & \text{if } \dfrac{\partial z_\circ}{\partial x}=0 \end{cases}$$

$$(59)\qquad n=Eq_P\cdot\frac{\partial z}{\partial e}\left(\frac{\partial e}{\partial b}\right)_{\bar{A}}.$$

Substituting (55) into (53) and noting c can be arbitrary if $\frac{\partial z_\circ}{\partial x}=0$, the optimal pair (b, c) is given by:

$$(60)\qquad (1-b)(m+n)=(\theta_A-\theta_P)(Ez)$$

$$(61)\qquad c=k(1-b)$$

where $k=0$ if $\dfrac{\partial z_\circ}{\partial x}=0$

$$k=\frac{Eq_P\left(\dfrac{\partial z}{\partial x}+\dfrac{\partial z}{\partial e}\cdot\dfrac{\partial e}{\partial x}\right)}{\dfrac{\partial z_\circ}{\partial x}} \quad \text{if } \frac{\partial z_\circ}{\partial x}>0.$$

These two equations (60) and (61) are the fundamental equations of the optimal linear incentive system. Here, θ_P and θ_A measure, in a sense, the degree of concavity of P and A. θ_P and θ_A are zero if P and A are linear in w and v. m and n are the weighted averages of the responsiveness of output through a risk decision and an effort decision to a change in b. k is a measure of the relative responsiveness of output z to a change in x, in comparison with the responsiveness of the goal.

An implication of (55) is significant. It means that, if $\frac{\partial z_\circ}{\partial x}>0$, i.e., if $z_\circ$ is related to x, the risk decision that the agent chooses satisfies both $\frac{\partial}{\partial x}EP=0$ and $\frac{\partial}{\partial x}EA=0$ at the same time. Unanimity on x decision is guaranteed by a suitable design of an incentive system. The risk decision by the agent can be influenced in any way the principal wants by changing c. Thus, when $\frac{\partial z_\circ}{\partial x}>0$, the optimal b is determined

by n only, i.e., the effort decision effect of b. The risk decision effect of b plays no part since influencing the agent's risk decision is handled by c.

If we define $m'=\frac{b}{E(z)}m$ and $n'=\frac{b}{E(z)}n$, we may call them the average compensated elasticity of the principal's utility through a risk decision and an effort decision respectively. For example,

$$n'=\frac{b}{EP_w\cdot(1-b)E(z)}EP_w\cdot\frac{\partial w}{\partial e}\cdot\left(\frac{\partial e}{\partial b}\right)_{\bar{A}}$$

Although $EP_w\cdot(1-b)E(z)$ is not exactly equal to EP and only a rough approximation, n is akin to "average elasticity". We may also interpret m' and n' as the weighted average compensated elasticity of output (z), the weights being q_P, through a risk decision and an effort decision respectively. To see this, we note

$$n'=\frac{b}{E(z)}Eq_P\cdot\frac{\partial z}{\partial e}\cdot\left(\frac{\partial e}{\partial b}\right)_{\bar{A}}$$

and similarly for m'.

Using m' and n', we can transform (60) further into:

$$b=\frac{m'+n'}{m'+n'+\theta_A-\theta_P} \tag{62}$$

This result generalizes the result obtained by Stiglitz [11] in that a risk-averse principal is considered and both risk and effort decisions are incorporated sequentially in a general setting. The sharing parameter b depends on the weighted average compensated elasticity of output and the difference in the degrees of concavity of A and P.

In (62), θ_A and θ_P play a crucial role. One of their intuitive interpretations is the negative of "normalized covariance" between marginal utility and output (z). It is easy to see $\theta_A=-\frac{cov(z, A_v)}{E(z)EA_v}$, where cov(,) means covariace. If $0<b<1$ as most likely, θ_A and θ_P are always non-negative because $A_{vv}\leq 0$ and $P_{ww}\leq 0$. If $m'+n'>0$ (as is likely), greater θ_A or smaller θ_P or greater $\theta_A-\theta_P$ all lead to smaller b. If we may interpret θ as a degree of risk aversion (we will show this below), (62) is a quite plausible result. Furthermore, if $\theta_A>\theta_P$ as is often believed (the agent is more risk averse than the principal), the greater $m'+n'$ (i.e., incentive effects of b), the greater b.

Interpreting these results, we should be cautious because θ_A and θ_P depend on the magnitude of b through A_v and P_w (because of b's direct effect on v). A simplistic comparative static from (62) may be misleading because both sides of (62) depend on b. In order to avoid this, we need further characterization of θ_A and θ_P using more familiar concepts like absolute risk aversion, etc. To do so, let us assume that the following first-order approximation of marginal utility of income is reasonable: (the case for P is similar and omitted)

$$A_v=\bar{A}_v+\bar{A}_{vv}(v-\bar{v})+\bar{A}_{ve}(e-\bar{e}) \tag{63}$$

where $\bar{v}=E(v)$, $\bar{e}=E(e)$, and $\bar{A}_v$, etc., are the values of the functions evaluated at $v=\bar{v}$, $e=\bar{e}$. If we further assume $\bar{A}_{ve}=0$ or the covariance between e and z is small, we obtain

$$\theta_A = \bar{R}_A \cdot \frac{\sigma^2}{E(z)} b = \frac{\bar{R}_A \cdot \sigma_v^2}{bE(z)} \tag{64}$$

where R_A is R_A evaluated at $\bar{v}$, $\bar{e}$ and σ^2 is the variance of z. σ_v^2 is the variance of v. Similarly,

$$\theta_P = \bar{R}_P \cdot \frac{\sigma^2}{E(z)}(1-b) = \frac{\bar{R}_P \cdot \sigma_w^2}{(1-b)E(z)} \tag{65}$$

where σ_w^2 is the variance of w.

From these approximation results, the meaning of θ becomes clearer. Since $1/2\ \bar{R}_A \sigma_v^2$ is the Arrow-Pratt risk premium, θ_A and θ_P may be called "normalized risk premiums" (not quite the same as proportional risk premium). (64) and (65) make the dependence of θ on b much more explicit. It is interesting to note at this stage:

PROPOSITION 3. Assume $E(z)>0$ and $A_{ve}=0$.

(i) if $m+n>0$, $0 \leqq b \leqq 1$ and $\theta_A \geqq \theta_P$

(ii) if $m+n=0$, $\theta_A = \theta_P$

(iii) if $m+n<0$, either $b \leqq 1$ and $\theta_A \leqq \theta_P$,
or $b \geqq 1$ and $\theta_A \geqq \theta_P$

Proof.

(i) If $m+n>0$, we see from (60) that $1-b$ and $\theta_A - \theta_P$ have to be of the same sign. If $b>1$, then $\theta_A \geqq 0$ and $\theta_P \leqq 0$ because θ is the negative of the normalized covariance. But this contradicts (60) because $b>1$ implies $\theta_A - \theta_P < 0$. If $b<0$, then $\theta_A \leqq 0$ and $\theta_P \geqq 0$. Again this contradicts (60) because $\theta_A - \theta_P > 0$ if $b<0$.

The proof of (ii) and (iii) are obvious.

The case $m+n>0$ would be most likely and interesting. The optimal sharing parameter is between 1 and 0 and it is determined in such a way that the agent's normalized risk premium is greater than (or equal to) that of the principal. In other words, the optimal incentive system that the principal selects always makes the agent "no less risk averse" than the principal. Therefore, if the agent is risk-neutral and the principal is risk-averse, the only way to make the agent "no less risk averse" is to make the principal's risk premium zero. Then, it is necessary to free himself from risk, i.e., $b=1$. This easily follows from (60) by noting $\theta_A=0$ and $-\theta_P \leqq 0$ (and $\theta_P=0$ if $b=1$). When $m+n=0$ (i.e., b has no marginal effects through x and e), b is selected so that the normalized risk premium for both people are the same. (This occurs, as we will see below, when $b=\frac{\bar{R}_A}{\bar{R}_A+\bar{R}_P}$).

From (64) and (65), we can rearrange (60) and get: (From now on, we omit — on top of R_A and R_P for notational simplicity.)

$$\{m+n+\sigma^2(R_A+R_P)\}b = m+n+\sigma^2 R_P \tag{66}$$

Solving for b, unless $m+n=-\sigma^2(R_A+R_P)$,

$$(67) \qquad b = \frac{m+n+\sigma^2 R_P}{m+n+\sigma^2(R_A+R_P)}$$

Thus, we obtain the following proposition. The proof is simple and thus omitted.

PROPOSITION 4. Let $b^*=\dfrac{R_P}{R_A+R_P}$

(i)	If $m+n>0$	$1\geq b>b^*$
(ii)	If $m+n=0$	$b=b^*$
(iii)	If $0>m+n>-\sigma^2 R_P$	$b^*>b>0$
(iv)	If $m+n=-\sigma^2 R_P$	$b=0$
(v)	If $-\sigma^2 R_P>m+n>-\sigma^2(R_A+R_P)$	$0>b$
(vi)	If $m+n=-\sigma^2(R_A+R_P)$	no b
(vii)	If $m+n<-\sigma^2(R_A+R_P)$	$b>1$
(viii)	$b=1$ when $R_A=0$	

Depending on the magnitude of $m+n$, this proposition shows how the optimal sharing parameter changes. The result is much stronger than Proposition 3, although at the cost of the approximation assumption (63). It is interesting to note that for a wide range of $m+n$, b is between 0 and 1 and $b>0$ even if $R_P=0$ as long as $m+n>0$. A risk-neutral principal sets b at a positive level to secure the incentive effects. When $R_A=0$, the optimal b is unity, meaning that the agent absorbs all the risk. Even if $R_P=0$ at the same time, the principal still prefers to let the agent pay him "a fixed fee" (a would be negative if $b=1$) because of the effort effect.

Another interesting observation from Proposition 4 is that for $m+n>0$, b has a lower limit, $\dfrac{R_P}{R_A+R_P}$, which is the optimal b when $m+n=0$. One of the cases for which $m+n=0$ is when there is no risk and effort decisions in the situation in the first place. In such a case, the role of a incentive system is purely distributional. It determines how the uncertain output is distributed among two people. The optimal b in such a case is easy to interpret as a risk-sharing parameter. In fact, Stiglitz [11] has obtained $b=b^*$ when there is no agent's decisions (x or e) involved.

The solution $b=b^*$ is also obtained in Wilson [14] and Ross [9] in a decision situation with only x (no e), using Pareto optimality as mentioned above, as the definition of optimality. In our framework, $m+n=0$ only when $m=0$ if there is no effort decision. From (55) and (58) it then follows that $m=0$ when $\dfrac{\partial}{\partial x}EP=0$, i.e., when there is unaminity on x between the agent and the principal, whether z_0 is meaningfully included in the incentive system $\left(\dfrac{\partial z_0}{\partial x}\neq 0\right)$ or not $\left(\dfrac{\partial z_0}{\partial x}=0\right)$. When unaminity exists, it is obvious that the solution of (46) (the principal's optimization) gives the Pareto optimal incentive system. If people agree on the decision to be made, the only remaining problem of incentive system design is that of distribution of output among them. The pareto

optimal risk-sharing from a distributional point of view (b^*) determines the optimal incentive system.

Following this line of reasoning, we may then interpret $b-b^*$ as a "decision premium" or "incentive premium" of the optimal sharing parameter. This is an increment (or decrement) of b over b^* (distributionally optimal b; let us call this the risk-sharing premium) which the principal wants to have in order to influence the agent's decision. If $m+n>0$, we may roughly argue that incentive premium is positive because the incentive effect (both risk and effort) of increasing b is still positive. It is just the opposite when $m+n<0$. The principal thinks that $b=b^*$ is too high a level of b and it has negative incentive effect, either by making the agent exert less effort because of too much income already, or by making the agent's risk decision too conservative from the principal's point of view.

We can further decompose the incentive premium into the effort incentive premium and the risk incentive premium, using (67). One way to do this is:

$$b-b^*=(1-b^*)\left\{\frac{n}{\sigma^2(R_A+R_P)+n}+\left(1-\frac{n}{\sigma^2(R_A+R_P)+n}\right)\frac{m}{\sigma^2(R_A+R_P)+m+n}\right\} \tag{68}$$

or

$$b=b^*+(1-b^*)\frac{n}{\sigma^2(R_A+R_P)+n}+(1-b^*)\left(1-\frac{n}{\sigma^2(R_A+R_P)+n}\right)\times\frac{m}{\sigma^2(R_A+R_P)+m+n} \tag{69}$$

The second term in (69) is the effort incentive premium and the third term the risk incentive premium. Assuming $0<b<1$, the effort incentive premium is positive when $n>0$ and the risk incentive premium is positive or negative depending on $m>0$ or $m<0$. From (58), we see that the risk incentive premium is zero when $\frac{\partial z_0}{\partial x}>0$. That is, the risk decision incentive is handled entirely by c, as was already mentioned.

The equation (68) or (69) indicates that

the optimal b=the risk-sharing premium+the effort incentive premium
+the risk incentive premium

and this decomposition in a sense summarizes the history of the theoretical discussions of the optimal linear incentive system. Wilson [14] first considered the optimal risk sharing arrangement and found the risk-sharing premium. Stiglitz [11] then discussed effort, and actually discussed an amalgam of two incentive premium together with the risk-sharing premium. Ross [9] discussed the risk incentive premium without an effort decision.

From the results of the previous section on $\frac{\partial x}{\partial b}$ and $\frac{\partial e}{\partial b}$, it is perhaps not so unreasonable to assume $n>0$ and $m\leqq 0$ $\left(m=0 \text{ when } \frac{\partial z_0}{\partial x}>0\right)$. Then, a possible verbal argument that we may offer as a paraphraise of (69) on the way the optimal b is determined is the following: First, the risk-sharing premium is determined from the distributional point of view. Then, from the decision influence point of view, the effort

incentive premium is added to the risk-sharing premium. But the resulting b is often too great and causes a conservative risk decision by the agent and thus the risk incentive premium is "deducted" to strike the optimal balance among the three kinds of effects. If we fail to recognize the existence of the agent's risk decision, the "optimal" b would overestimate the truly optimal b. If we fail to recognize an effort decision, the "optimal" b would be below the true optimum.

It would be worthy if we can derive some comparative static results on b and c. To do so, it would be better to use m' and n' (elasticities) rather than m and n (derivatives) in describing b (and c). From (62), we obtain

$$\sigma^2(R_A+R_P)b\left(b-\frac{R_P}{R_A+R_P}\right)=E(z)(m+n)(1-b)$$

Then, it is easy to see (proof omitted):

PROPOSITION 5. *Ceteris paribus*, the optimal b would be greater when m', n', $E(z)$, or R_P increases and when σ^2 or R_A decreases.

The implications of this proposition is quite reasonable. When the output responsiveness is greater (m', n'), then the agent would share more in the output. When the agent is more risk-averse or the environment more uncertain or the principal is less risk-averse, the agent shares less in the uncertain output.

As was mentioned several times, the principal can induce unanimity or a risk decision by changing c. Let us briefly investigate how important this goal factor is in the optimal incentive system. First, from (61) and (67)

$$\frac{c}{b+c}=\frac{kR_A}{kR_A+R_P+\frac{n}{\sigma^2}}$$

Thus:

PROPOSITIONS 6. Assume $0<b<1$ and $k>0$. The relative importance of the goal, $\frac{c}{b+c}$, increases, *ceteris paribus*, when R_A or k increases or when R_P or n decreases. If $n>0$, $\frac{c}{b+c}$ increases when σ^2 increases.

We can see that when the divergence between risk aversions of the principal and the agent is great (the agent being more risk-averse), the importance of c as a risk-decision influencing device is great. When n is large, i.e., output responsiveness of b through effort is large, the goal is less important. To see the implication of k more concretely, let us suppose that the agent reports the goal as a linear function of $E(z)$, i.e., $z_0=\alpha E(z)+\beta$. Then $k=\frac{1}{\alpha}$ measures the degree of biased reporting. The larger k, the more downward bias the goal has (i.e., $z_0<E(z)$) unless β is a significant positive number. Then, Proposition 6 implies that the gerater the downward bias, the more the importance of c. In order to maintain c's influence on a risk decision so that unanimity on x can be secured, it is necessary to make c larger when there is some downward bias in goal reporting. Otherwise, b's effect on conservative risk decision cannot be offset.

V. *Concluding Remarks*

In this paper, we have been concerned with the optimal determination of a linear goal-based incentive system under a two-stage sequential decision process by the agent. When the goal is related to the agent's risk decision and if the principal knows this relationship, the principal is found to use the goal incentive parameter (c) to achieve unanimity in a risk decision. In fact, z_0 can be anything as long as it is related to x in a non-random fashion. The result here implies that as long as a variable is available which is related to x, it is better for the principal to use this variable in the optimal incentive system design. By incorporating such a variable into the incentive system (i.e., $c \neq 0$), it is likely that the principal can rely on b more (i.e., can make b greater) seeking more effort incentive effect without the fear of letting the agent share too much risk, thus causing too conservative a risk decision. (Obviously, this is true when $m>0$.) In other words, the inclusion of the goal level may have a risk-encouraging effect.

As mentioned in Section 2, the goal is often included in an incentive system as a deviation $(z-z_0)$, as in (7). We can derive some implications of our analysis on the optimality of a deviation-based incentive system of the form (7). First, it is unlikely that $b'=c'=0$ and $r>0$ is optimal in (7) as long as $\frac{\partial z_0}{\partial x}>0$. That is, the incentive system based solely on the deviation (and a fixed payment, a) is unlikely to be optimal. For $b'=c'=0$ and $r>0$ to be optimal, we need $b=-c>0$ and this implies

$$n+\sigma^2(R_A+R_P)=(1-k)\sigma^2R_A$$

and either

$$k<0,\ r<1,\ n+\sigma^2R_P>0$$

or

$$k>1,\ r>1,\ n+\sigma^2(R_P+R_A)<0.$$

These are hardly the conditions we would expect to hold in general.

The second implication of our analysis is that for $r>0$ to be optimal it is likely necessary to have $c'>0$, although b' can be zero. If $c'>0$ and $b'=0$, the incentive system becomes $v=a+rz+(c'-r)z_0$. By suitably selecting c' and r, the optimal incentive system can be obtained. In other words, when the deviation becomes a sigificant factor in the incentive system, the goal level itself should often receive a considerable attention. Thirdly, Whether $b'=c'=0$ or $b'=0$, $c'>0$, the optimal r is smaller than 1 for a wide range of cases (because $r=b$ in these cases). This is obvious from Proposition 4. Too great dependence on the deviation $(r>1)$ is unlikely to be optimal even though the deviation may be a small number in magnitude.[12]

By assuming the simultaneous existence of a risk decision and an effort decision, we could see that the optimal sharing parameter b is composed with three parts: risk-sharing premium, effort incentive premium and risk incentive premium. The sum of the latter two may be called incentive premium or decision premium. If we ignore the

[12] These implications on a deviation-based incentive system are based on a linear incentive function. Once we admit non-linearity in treating the deviation in an incentive system, the conclusions may change.

existence of either one of two decisions by the agent, we misjudge the magnitude of incentive premium. When an effort decision is ignored, it is likely to result in smaller b than the true optimum. Ignoring a risk decision likely leads to too great a sharing parameter.

The existence of an incentive premium is in a sense a sign of non-cooperativeness of the agency relationship. If two people are unanimous on both a risk decision and an effort decision, there is no need to have an incentive premium. The only problem of an incentive system is how to divide the output among them. The incentive premium is like unavoidable cost of delegation of decisions from the principal to the agent. In order to minimize the cost of delegating decisions, the delegator of decisions often sets up institutional arrangements to influence the decisions by the delegatee. One of them is the incentive system we discussed in this paper. A monitoring and screening system of the delegatee's behavior, ability and so on is another example of such institutional arrangements. In fact, a deviation-based system we touched in this paper seems to work as a monitoring system as well as an incentive system in reality. The rational explanation of these institutional arrangements in a general agency relationship (and the organization in particular) has begun only recently and has been the basic motivation of the present paper.

REFERENCES

[1] Aoki, M., "Risk-Sharing Employment Contracts vs. the Auction Market," mimeo, Kyoto Institute of Economic Research, Kyoto University, March 1978.

[2] Arrow, K.J., *Essays in the Theory of Risk-Bearing*, North-Holland, 1974.

[3] Atkins, A.A., "Standard Setting in an Agency," *Management Science*, September 1978, pp. 1351-1361.

[4] Demski, J.S., and G.A. Feltham, "Economic Incentives in Budgetary Control Systems," *Accounting Review*, April 1978, pp. 336-359.

[5] Harris, M., and A. Raviv, "Some results on Incentive Contracts with Applications to Education and Employment, Health Insurance, and Law Enforcement," *American Economic Review*, March 1978, pp. 20-30.

[6] Itami, H., "Analysis of Implied Risk-taking Behavior under a Goal-Based Incentive Scheme," *Management Science*, October 1976, pp. 183-197.

[7] Mirrlees, J.A., "The Optimal Structure of Incentives and Authority within an Organization," *Bell Journal of Economics*, Spring 1976, pp. 105-131.

[8] Ross, S.A., "The Economic Theory of Agency: The Principal's Problem," *American Economic Review*, May 1973, pp. 134-139.

[9] ———, "On the Economic Theory of Agency and The Principle of Similarity," in *Essays on Economic Behavior under Uncertainty* (M.S.Balch, et. al. ed.), North-Holland, 1974, pp. 215-237.

[10] Stiglitz, J.E., "The Effects of Income, Wealth, and Capital Gains Taxation on Risk-Taking," *Quarterly Journal of Economics*, May 1969, pp. 263-283.

[11] ———, "Incentives and Risk Sharing in Sharecropping," *Review of Economic Studies*, April 1974, pp. 219-255.

[12] ———, "Incentives, Risk, and Information: Notes towards a Theory of Hierarchy," *Bell Journal of Economics*, Autumn 1975, pp. 552-579.

[13] Weitzman, M.L., "The New Soviet Incentive Model," *Bell Journal of Economics*, Spring 1976, pp. 251-257.

[14] Wilson, R., "The Structure of Incentives for Decentralization under Uncertainty," in *La Decision* (Edition du Centre National de la Recherche Scientifique), Paris, 1969, pp. 287-397.