Commitment Problem, Optimal Incentive Schemes, and Relational Contracts in Agency with Bilateral Moral Hazard

著者	SUZUKI Yutaka
出版者	Institute of Comparative Economic Studies, Hosei University
journal or publication title	Journal of International Economic Studies
volume	21
page range	103-124
year	2007-03
URL	http://doi.org/10.15002/00002049

Journal of International Economic Studies (2007), No.21, 103–124

Commitment Problem, Optimal Incentive Schemes, and Relational Contracts in Agency with Bilateral Moral Hazard

Yutaka Suzuki
Faculty of Economics, Hosei University

Abstract

In a bilateral moral hazard framework, where the principal is also a productive agent, the requirements of both the agent's and the principal's incentive provisions should be satisfied in designing optimal incentive contracts. In a static framework, only the second best is obtainable if the incentive contract is based only on the total output. One example of this is the simple linear sharing rule that is often observed. Next, it is shown that in a repeated game version, such a commitment problem could be solved, and a first best outcome could be achieved through both parties taking trigger strategies that depend on a public signal. We give an interpretation in the viewpoint of the 'reputation' mechanism, and a qualitative characterization on the optimal solution induced in equilibrium for *all* possible discount factors. Finally, some applications for corporate governance are presented.

Keywords: Bilateral Moral Hazard, Team Production, Commitment Problem, Linear Contracts, Relational Contracts, Reputation and Self Enforcement, Corporate Governance

JEL Classification C72, D82, L23

1. Introduction

In the traditional principal-agent problem, it is assumed that the principal delegates the agent to act on behalf of the principal, which cannot be observed by the principal. This situation brings about an incentive problem. In order to mitigate it, the principal delegates the agent all the functions as productive inputs and designs an incentive contract that is based on common observables[1], such as outputs, realized costs, and profits, which are correlated with the agent''s hidden action choice.

It is well-known that many incentive contracts give a socially efficient (first best) outcome when the principal and the agent are risk-neutral. When both parties are risk-neutral, risk sharing is not an issue. Only the incentive provision for the agent''s action choice is an issue in designing a sharing rule. One simple example that generates the first best solution is giving the agent all the authority and risk of production and simply charging a fixed rent.[2] When the agent is risk averse, not only the incentive provision but also the optimal risk sharing between the principal and the agent is an important issue in designing an optimal incentive

[1] Through the model, we assume that these common observables can be verified also to the court without cost. In other words, there exists verifiable information, on which the principal can write an incentive contract.

[2] This incentive contract performs well unless the agent's limited liability matters.

contract. There is a trade-off between two alignments. Thus, the socially efficient (first best) outcome cannot be obtained in this case. This is also well-known.

Rather, in this model, we shall assume that the principal also participates in the production process and invests her own effort that is not observable to the agent. The assumption that the principal accesses the production process looks uncommon but is more realistic. Usually, the principal provides the workers with production machinery, the quality of which is related to the principals'effort input and sometimes not observable to the workers. The principal's advertising investment or lobbying effort is another example. Advertising investment is not usually observable to the worker. Further, even though the advertisement itself may be observable, it may not be verifiable without costs. In this framework, not only the incentive provision for the agent''s action choice but also the principal's credible commitment on her effort level is an issue in designing an incentive scheme.

This framework is similar to the literature that discuss the moral-hazard problem in a team production environment in the sense that *after* the principal sets up the incentive contract as a Stackelberg leader, she herself participates in the production as a team member. So, the principal here can be viewed as a principal agent. In such situations, in the case where only the contract based on the total output is feasible, there will be a trade-off in designing an incentive contract whereby the principal should provide the agent with the incentive to take an optimal action and at the same time she should *credibly commit herself* to take her own best action. Giving incentive to the agent to work hard *always hurts* her own credible commitment to some extent. Thus, only the second best efficiency is attainable. In Section2, we show that there always exists a simple linear sharing rule that obtains such a second best outcome. Also, it is discussed that each party's sharing proportion under this sharing rule crucially depends upon each party's marginal contribution to total production.

In Section 3, with a simple example, it is shown that the agent's sharing proportion increases as his relative contribution to production gets larger. When the agent's contribution to total production gets more important, more stress should be given on the agent's action in the action combination. Thus, the agent's sharing portion should increase to impose more weight on the agent's action incentive. However, the essential point is discussed that the additional requirement for the principal's credible commitment about her optimal effort restricts her degree of freedom in designing a contract.[3]

In Section 4, we introduce the idea of *relational contracts*-contracts that are *self-enforced through repeated interaction* between the principal and the agent-rather than pure court-enforced (formal) contract. We show that in a repeated game version of the model, the commitment problem mentioned above could be solved, and a first best outcome could be achieved through both parties taking *trigger strategies* that depend on a public (but unverifiable) signal. Then, we give an interpretation in the viewpoint of the 'reputation' mechanism, and a qualitative characterization about the optimal solution induced in equilibrium for all possible ranges of discount factor. This approach can be interpreted as analyzing how the implementable incentives change under the connection with the *formal contract and the relational contract*, for all discount factors.[4] Then, in Section 5, we give an extension of the model, based on a blend of the ideas of "identification of deviations" by Legros-Matthews (1993) and "contingent control shift" by Aghion-Bolton (1992), and last interpret our result in a context of 'multiple equilibria' and discuss the renegotiation-proofness problem in repeated

[3] This structure is applied also to the dynamic framework, by replacing the static first order condition with the dynamic self-enforcing constraint.

[4] Like Macleod-Malcomson (1989), there is a complementarity between formal and relational contracts in our model.

games and its implication for recent corporate restructuring.

In Section 6, we conclude the paper.

1.1 Related Literature

Situations where the principal may also have to extend unobserved effort are uncommon in the literature, but seem to be more realistic in empirical observation. Bhattacharya and Lafontaine (1995) consider this case in the context of franchise, where the upstream and downstream parties share the rents which their relationship generates. Indeed, the former part of this paper generalizes their model by assuming an output function with a much less restrictive form of separability, and gives a clear proof and intuition.[5] However, from the viewpoint of practice, there is another important element in franchising relationships: continuation of the relationship, which is also common to the employment relationship within firms. The role that *relational contracts*—contracts that are *self-enforced through repeated interaction* between the parties—play seems to be important, as well as pure court-enforced (*formal*) contracts. Hence, unlike Bhattacharya and Lafontaine (1995), we combine a relational contract (self-enforcing contract) perspective with an agency model with bilateral moral hazard. The second aspect of our model is closely related to the recent contract theoretic/game theoretic relational contracting literature. Levin (2003) studies *self-enforced incentive contracts* under imperfect observability, using a model of repeated agency.[6] Rayo (2001) extends Levin's model to multi-agent situations, and shows the endogenization of the principal- multi agent structure, in a repeated agency setting. The differences from our model are that they do not analyze a bilateral moral hazard situation and in their model the threat point payoff when the cooperation is broken in the continual relationship is just an *exogenous* default payoff, whereas in our model it is the repetition of the static second best equilibrium (so-called 'one-shot Nash reversion' in the repeated game literature). Baker, Gibbons and Murphy (1994, 2002) examine an interaction between asset ownership (*who owns the asset*) and relational contracts both within (*employment*) and between firms (*outsourcing*). Our framework belongs to the spot and relational employment relationships within firms (the principal owns the asset, e.g. production machinery). However, they do not analyze *the bilateral moral hazard situation* the way our model does.

2 The Model

A risk-neutral principal owns a firm. She hires a risk-neutral agent and designs a wage contract. $a \in A = [0, \infty)$ and $e \in E = [0, \infty)$ denote the efforts provided by the agent and the principal, respectively. $Q: A \times E \to R$ represents a composite function of both efforts, i.e. $q = Q(a, e)$. $Q(a, e)$ is continuous a and e differentiable in and. It is assumed that both efforts are productive, i.e., $Q_a > 0$ and $Q_e > 0$, where subscripts indicate partial derivatives. The output y that is commonly observable (and verifiable) is a function of the effort composition q and

[5] Nandeibam (2002) examines almost the same problem in a more general framework with n-agents. However he does not mention the contract offer from the principal to the agent and just *assumes* the positive action pairs in equilibrium. On the other hand, we *explicitly derive* how the formal contract offered by the principal implements the second best positive actions in the static equilibrium (Proposition 4), and moreover investigate the interaction (complementary role) between the formal and the relational contracts, and characterize the optimal action levels in that case for *all* discount factors (Proposition 5 and its corollary).

[6] Bolton and Dewatripont (2005) give a survey of the optimal relational contracts analyzed by Macleod-Malcomson (1989) and Levin (2003).

the state of nature $\theta \in \Theta$, that is, $y = Y\left(Q\left(a,e\right),\theta\right)$, where $Y: Q \times \Theta \to R$.[7] We can suppres θ and consider y directly as a random variable, parameterized by the effort composition q. Then, $F\left(y \middle| q\left(a,e\right)\right)$ denotes the output distribution conditional on the effort composition, and $f\left(y \middle| q\left(a,e\right)\right)$ denotes the output probability density function.

The agent's wage scheme w is based on the observable (verifiable) y. $\Psi\left(e\right)$ and $C\left(a\right)$ denote the disutilities (costs) of the principal's and the agent's efforts, respectively, with $\Psi'\left(e\right) \geq 0, \Psi''\left(e\right) > 0$ and $C'\left(a\right) \geq 0, C''\left(a\right) > 0$. Further, it is assumed that $\Psi\left(0\right) = C\left(0\right) = 0$ and $\Psi'\left(0\right) = C'\left(0\right) = 0, \Psi'\left(\infty\right) = C'\left(\infty\right) = \infty$.

Assumption 1
f is twice differentiable in q.

Assumption 2
$F\left(y \middle| q_1\right) \geq F\left(y \middle| q_2\right), \forall y, q_1 \leq q_2$

Assumption 3
$F\left(y \middle| Q\left(a,e\right)\right)$ is convex in a and e.

Assumption 1 is needed to guarantee the existence of the agent's and the principal's optimal effort choices. Assumption2 denotes that the effort composition q is productive in the sense of first order stochastic dominance.[8] Assumption 3 is given for the uniqueness of the optimal action combination $\left(a,e\right)$ and means the decreasing marginal productivity of a and e in a probability sense.[9]

2.1 The First Best

Let us temporarily assume that the principal can observe the agent's effort choice directly .The principal's optimization problem is

$$\underset{w(y),a,e}{Max} \int \left(y - w\left(y\right)\right) f\left(y \middle| Q\left(a,e\right)\right) dy - \Psi\left(e\right)$$

$$s.t \int w(y) f\left(y \middle| Q\left(a,e\right)\right) dy - C\left(a\right) \geq \overline{U}$$

The constraint is the agent's participation constraint that requires his expected utility obtained from this production process not less than that of other alternatives denoted by $\overline{U}$. Since the principal can observe the agent's action choice, a forcing contract is possible.

It is obvious that at the optimum

$$\int w\left(y\right) f\left(y \middle| Q\left(a,e\right)\right) dy - C\left(a\right) = \overline{U} \tag{1}$$

Thus, the principal's optimizing problem is

[7] Note that q and θ independently affect.

[8] For a simple explanation of first order stochastic dominance, see, Bolton and Dewatripont (2005).

[9] Holmstrom (1982) used this assumption (p 328). This assumption is stronger than the assumption $E\left[y \middle| Q\left(a,e\right)\right]$ is concave in a and e. In the no uncertainty case, it is enough for the uniqueness of optimal effort combination. He also remarks that this is not satisfied in some very natural cases on distribution functions.

$$\underset{w(y),a,e}{Max}\; E\left[y\middle|Q(a,e)\right] - \Psi(e) - C(a) - \overline{U}$$

where $E\left[y\middle|Q(a,e)\right] \equiv \int yf\left(y\middle|Q(a,e)\right)dy$

Assumption 3 and the convexity of $\Psi(e)$ and $C(a)$ guarantees the uniqueness of the optimal effort combination (a^*, e^*). At the optimum, the principal chooses her effort level e^* that satisfies $\frac{\partial E\left[y\middle|Q(a^*,e^*)\right]}{\partial e} = \Psi'(e^*)$ and forces the agent to choose a^*

such that $\frac{\partial E\left[y\middle|Q(a^*,e^*)\right]}{\partial a} = C'(a^*)$ where the primes denote the first derivatives. The only requirement imposed on $w^*(y)$ is to satisfy (1). One of the simple examples of $w^*(y)$ is a fixed fee, that is, $w^*(y) = \overline{U} + C(a^*), \forall y$. With a fixed wage scheme, the principal remains a residual *claimant* and obviously has an incentive to choose e^* .

2.2 The Second Best

Now assume that the principal cannot observe the agent's effort choice and also monitoring it is impossibly costly. A forcing contract is impossibly costly. All she can do is to give the agent incentive to choose an optimal action through a wage contract $w(y)$. In a traditional principal-agent model, where the agent is the only party who provides the effort inputs, it is well known that charging the agent a fixed rent and giving all the reminder of the output generates the first best solution when the agent is risk-neutral. However, when both the principal and the agent invest their own productive inputs together, such a wage contract cannot implement the first best outcome, since the principal should not only give the agent incentive to do optimally but also commit herself to do optimally through the wage contract in this bilateral moral hazard setting. Under the contract that is strategically equivalent to selling the firm to the agent, the principal cannot convince the agent that she does her best because she (the principal) only gets *a fixed rent*. Thus, in this bilateral moral hazard setting, the principal's designing of an incentive scheme is substantially constrained by two incentive problems. One is the agent's incentive constraint and the other is her own incentive constraint. Hence, the principal's problem becomes

$$\underset{w(y),a,e}{Max} \int \left(y - w(y)\right) f\left(y\middle|Q(a,e)\right)dy - \Psi(e)$$

$$s.t(i)\; \int w(y) f\left(y\middle|Q(a,e)\right)dy - C(a) \geq \overline{U}$$

$$(ii)\; (a,e) \text{ satisfies}$$

$$a \in \arg\max \int w(y) f\left(y\middle|Q(a,e)\right)dy - C(a)$$

$$e \in \arg\max \int \left(y - w(y)\right) f\left(y\middle|Q(a,e)\right)dy - \Psi(e)$$

The second constraint requires that the effort combination (a,e) be a ("Static") Nash equilibrium that satisfies two incentive constraints, while the first one denotes the agent's participation constraint.

It is easily shown that at the (second best) optimum the participation constraint is also binding, i.e.,

$$\int w(y) f\left(y \middle| Q(a,e)\right) dy - C(a) = \overline{U} \tag{2}$$

Let us denote $W_{\overline{U}}$ as the set of $w(y)$ satisfying (2), then the principal's problem is to choose $w(y) \in W_{\overline{U}}$ that maximizes the following problem:

$$\begin{aligned}
&Max\ E\left[y \middle| Q(a,e)\right] - \Psi(e) - C(a) - \overline{U} \\
&w(y) \in W_{\overline{U}}, a, e \\
&\qquad s.t\ (a,e)\ satisfies \\
&\qquad\quad a \in \arg\max \int w(y) f\left(y \middle| Q(a,e)\right) dy - C(a) \\
&\qquad\quad e \in \arg\max \int \left(y - w(y)\right) f\left(y \middle| Q(a,e)\right) dy - \Psi(e)
\end{aligned}$$

Since there is no $w(y)$ in the maximand, we can easily conjecture that a specific contractual form of $w(y)$ does not matter in deciding the principal's expected utility only if it induces the same effort combination as a Nash equilibrium and in $W_{\overline{U}}$. Further, by Assumption 3 and, $\psi''(e) > 0, C''(a) > 0$, the incentive compatibility constraints that guarantee the Nash equilibrium action can be replaced by the first-order conditions:

$$\begin{aligned}
&(a,e)\ satisfies \\
&\int w(y) f_Q\left(y \middle| Q(a,e)\right) \frac{dQ}{da} dy - C'(a) = 0
\end{aligned} \tag{3}$$

and

$$\int \left(y - w(y)\right) f_Q\left(y \middle| Q(a,e)\right) \frac{dQ}{de} dy - \Psi'(e) = 0 \tag{4}$$

Let us denote C^0 is the set of (a^0, e^0) where there exists at least one wage contract $w(y) \in W_{\overline{U}}$ under which (a^0, e^0) is induced as a (Static) Nash equilibrium satisfying (3) and (4).

Lemma 1:

C^0 is not an empty set.

Proof:

Let $w^0(y) = \overline{U}, \forall y$. It is obvious that $w^0(y) \in W_{\overline{U}}$ since $C(0) = 0$. Since $w^0(y)$ is a fixed wage contract, it does not give the agent incentive to work hard, so the agent chooses $a = 0$. However the principal has full incentive to work hard under this contract, that is, $e = e^*(0)$, where $e*(0)$ denotes the principal's effort choice of full incentive when the agent chooses $a = 0$. Obviously, $(0, e^*(0))$ satisfies (3) and (4). Q.E.D

Another possible example is that $w^0(y) = y - k$ where $k = E\left[y \middle| Q\left(a^*(0), 0\right)\right] - \left[\overline{U} + C\left(a^*(0)\right)\right]$. Under this contract, the agent has full incentive to work hard, that is, $a = a^*(0)$, where $a^*(0)$ denotes the full incentive of effort choice when the principal chooses $e = 0$, but since the principal is given a fixed rent k , he does not have any incentive to work hard, that is, $e = 0$. Since $w^0(y) \in W_{\overline{U}}$ and $\left(a^*(0), 0\right)$ satisfies (3) and (4), $\left(a^*(0), 0\right)$ is a Nash equilibrium action combination under wage contract $w^0(y)$.

Lemma 2:

C^0 is a bounded set.

Proof:
To see C^0 is a bounded set, it is enough to show that $0 \le a^0 \le a^*\left(e^0\right) < \infty$ and $0 \le e^0 \le e^*\left(a^0\right) < \infty$, $\forall\left(a^0, e^0\right) \in C^0$,where $a^*\left(e^0\right)$ denotes the agent's full incentive action when the principal can credibly commit herself to choose e^0 and similarly $e^*\left(a^0\right)$ denotes the principal's full incentive action.

(i) $0 \le a^0 \le a^*\left(e^*\right) < \infty$

Since $a^*\left(e^0\right)$ is the agent's full incentive action given e^0, at $a^*\left(e^0\right)$

$$E_Q\left[y \middle| Q\left(a^*(e^0), e^0\right)\right] \frac{\partial Q\left(a^*(e^0), e^0\right)}{\partial a} = C'\left(a^*\left(e^0\right)\right)$$

where $E_Q \equiv \dfrac{dE\left[y|Q\right]}{dQ}$. It is obvious that $a^*\left(e^0\right) < \infty$ since $E\left[y|Q\right]$ is concave in a and e $C'\left(\infty\right) = \infty$. However, when the principal cannot credibly commit e^0 , then she can only commit herself to choose e^0 through the wage contract that satisfies (4). Using (3), (4) reduces to

$$E_Q\left[y \middle| Q\left(a^0, e^0\right)\right] \frac{\partial Q\left(a^0, e^0\right)}{\partial e} - C'\left(a^0\right) \frac{\dfrac{\partial Q\left(a^0, e^0\right)}{\partial e}}{\dfrac{\partial Q\left(a^0, e^0\right)}{\partial a}} = \Psi'\left(e^0\right)$$

$$\Leftrightarrow E_Q\left[y \middle| Q\left(a^0, e^0\right)\right] \frac{\partial Q\left(a^0, e^0\right)}{\partial a} - C'\left(a^0\right) = \frac{\Psi'\left(e^0\right)}{\dfrac{\partial Q\left(a^0, e^0\right)}{\partial e}}$$

Since, $\Psi'\left(e^0\right) \ge 0, E_Q\left[y \middle| Q\left(a^0, e^0\right)\right] \dfrac{\partial Q\left(a^0, e^0\right)}{\partial a} - C'\left(a^0\right) \ge 0.$

By the concavity of $E\left(\bullet\right)$ and the convexity of $C\left(a\right)$ in a , we obtain $0 \le a^0 \le a^*\left(e^0\right) < \infty$.

(ii) $0 \le e^0 \le e^*\left(a^0\right) < \infty$

This can be shown just the same. Q.E.D

One thing to note from the proof of Lemma 2 is that any (static) contract that induces positive incentive on one side in Nash equilibrium suffers incentive loss on the other side.

Proposition 1:
The First best action $\left(a^*, e^*\right)$ is not in C^0

Proof
Given the principal's first best action e^* , let us assume that $w^0\left(y\right)$ gives the agent full incentive to choose a^* , that is, a^* satisfies

$$\int w^0\left(y\right) f_Q\left(y \middle| Q\left(a^*, e^*\right)\right) dy \frac{\partial Q\left(a^*, e^*\right)}{\partial a} = C'\left(a^*\right)$$

Then,

$$\int w^0\left(y\right) f_Q\left(y \middle| Q\left(a^*, e^*\right)\right) dy = \frac{C'\left(a^*\right)}{\dfrac{\partial Q\left(a^*, e^*\right)}{\partial a}} \tag{5}$$

If $\left(a^*, e^*\right)$ is a Nash equilibrium action under $w^0\left(y\right)$, by plugging (5) into (4), we should obtain

$$\frac{\partial E\left[y\middle|Q\left(a^*,e^*\right)\right]}{\partial e}-\frac{\dfrac{\partial Q\left(a^*,e^*\right)}{\partial e}}{\dfrac{\partial Q\left(a^*,e^*\right)}{\partial a}}C'\left(a^*\right)=\Psi'\left(e^*\right) \tag{6}$$

Since $a^*>0$, the second term on the left- hand side in (6) is strictly positive and (6) cannot be true because by the definition of e^*

$$\frac{\partial E\left[y\middle|Q\left(a^*,e^*\right)\right]}{\partial e}=\Psi'\left(e^*\right). \qquad \text{Q.E.D}$$

Proposition 1 says that in this moral hazard framework, there is no wage contract that implements the first best outcome. Intuitively, this bilateral moral hazard setting is *team production*, even though one of the team members is entitled to a principal who designs a contract and extracts all the surplus. In this point, the essence of Proposition 1 is equivalent to the idea founded in Holmstrom (1982), which deals with moral hazard problem in a team production environment. He shows that any sharing rule that satisfies *budget-balancing constraint* cannot achieve the first best result in team production.[10] Indeed, the budget balancing constraint $w\left(y\right)$ is also imposed here and the wage contract is a dividing rule of a given budget $y=Y\left(Q\left(a,e\right),\theta\right)$. What we can derive from (3) and (4) is that any wage contract that gives the agent strictly positive incentive also hurts the principal's ability of commitment to some extent if the budget balancing constraint is imposed.

Proposition 2:
If $\left(a^0,e^0\right)\in C^0$, then any contract $w^0\left(y\right)\in W_{\overline{U}}$ that satisfies either (3) or (4) satisfies the other constraint and finally guarantees $\left(a^0,e^0\right)$ as a Nash equilibrium action.

Proof:
If $\left(a^0,e^0\right)$ is in C^0 , then there exists at least one $w^0\left(y\right)\in W_{\overline{U}}$ such that

$$\int w^0\left(y\right)f_Q\left(y\middle|Q\left(a,e\right)\right)dy\frac{\partial Q\left(a^0,e^0\right)}{\partial a}-C'\left(a^0\right)=0 \tag{7}$$

and

$$\int\left(y-w^0\left(y\right)\right)f_Q\left(y\middle|Q\left(a^0,e^0\right)\right)dy\frac{\partial Q\left(a^0,e^0\right)}{\partial e}-\Psi'\left(e^0\right)=0 \tag{8}$$

Using (7), (8) changes to

$$\frac{\partial E\left[y\middle|Q\left(a^0,e^0\right)\right]}{\partial e}-\frac{\dfrac{\partial Q\left(a^0,e^0\right)}{\partial e}}{\dfrac{\partial Q\left(a^0,e^0\right)}{\partial a}}C'\left(a^0\right)=\Psi'\left(e^0\right) \tag{9}$$

and (9) is independent of $w^0\left(y\right)$. Q.E.D

As a result, Proposition 2 tells that the contractual form of the wage schemes does not matter in making $\left(a^0,e^0\right)$ as a (Static) Nash equilibrium if it induces either the principal e^0 with the

[10] Holmstrom (1982) breaks the budget balancing constraint and by giving group incentives with a penalty scheme achieves the first best outcome. In his model, the principal purely plays a role as a breaker of the budget balancing constraint and does not make any effort input, while in our model the principal is not the breaker of budget balancing and makes a productive effort. So, the function of the principal is totally different.

agent's given a^0, or the agent a^0 with the principal's given e^0, when (a^0, e^0) already belongs to C^0.[11][12]

Proposition3:

If (a^0, e^0) is already in C^0, then there always exists a linear wage contract $w(y) = ry + s \in W_{\overline{U}}$, where $0 \le r \le 1$, under which (a^0, e^0) is induced as a Nash equilibrium.

Proof:

Since (a^0, e^0) is already in C^0, to show that $w(y) = ry + s \in W_{\overline{U}}$ induces (a^0, e^0) as a Nash equilibrium action combination, it is enough to show that $w(y) = ry + s \in W_{\overline{U}}$ satisfies (7) by Proposition 2. In order for $w(y) = ry + s$ to be in $W_{\overline{U}}$

$$\int (ry + s) f\left(y \middle| Q(a^0, e^0)\right) dy - C(a^0) = \overline{U} \tag{10}$$

From (10), we obtain the fixed payment part of the wage

$$s = \overline{U} + C(a^0) - rE\left[y \middle| Q(a^0, e^0)\right] \tag{11}$$

Equation (11) tells that s will be determined when r is determined. Thus the only requirement to be shown is that there exists $0 \le r \le 1$ satisfying

$$\int (ry + s) f_Q\left(y \middle| Q(a^0, e^0)\right) dy \frac{\partial Q(a^0, e^0)}{\partial a} = C'(a^0) \tag{12}$$

that is, under $w(y) = ry + s$, the agent has an incentive to choose a^0 when the principal surely chooses e^0. (12) reduces to

$$r \frac{\partial E\left[y \middle| Q(a^0, e^0)\right]}{\partial a} = C'(a^0) \tag{13}$$

From Lemma 2, $0 \le a^0 \le a^*(e^0)$ and we can easily see from (13) that as r increases from zero to one, the agent's incentive compatible action choice also continuously increases from zero to $a^*(e^0)$ given the principal's e^0. Thus there exists r that induces the agent to take a^0 given the principal's e^0. Q.E.D

Since, as mentioned above, the principal will be indifferent with any contract only if it induces the same action combination as a Nash equilibrium, Proposition 3 tells that if $(a^0(*), e^0(*))$ is the second best action combination in C^0 that maximizes the principal's expected utility, then we always find a *linear wage contract* $w^0(y) = r^0 y + s^0$ with which such optimality can be achieved.[13]

---

[11] Proposition 2 critically hinges on the assumption that the principal and the agent are risk neutral. If the agent is risk averse, for example, the contractual form does matter in achieving optimality.

[12] In our model, even discontinuous group incentive schemes, *as long as they satisfy the budget balancing constraint*, cannot achieve the first best outcome in the static game framework, due to the result of Proposition 1 and 2.

[13] We are *not* restricting the principal's contract spaces to include only linear contracts. Rather, we allow the principal to choose arbitrary contracts but ask whether there is an equilibrium that is linear. Many other equilibria exist besides this, but the linear equilibrium has both the fascinating efficiency property and the simplicity such that it can often be observed in reality. Theoretically, this point is the same as the derivation of a *linear* Bayesian Nash equilibrium of the double auction, as in Chatterjee and Samuelson (1983).

Proposition 4:
At the second best optimum, $a^0(*) > 0$ and $e^0(*) > 0$.

Proof:
Assume that $e^0(*) = 0$ at the optimum, that is, the principal prefers to design a contract that induces $(a^*(0), 0)$ as a second best Nash equilibrium action. From Proposition 3, one such contract is $w(y) = ry + s$ where $r = 1$ and $s = \overline{U} + C(a^*(0)) - E\left[y \middle| Q(a^*(0), 0)\right]$ to guarantee $w(y) \in W_{\overline{U}}$. When the principal changes the contract by marginally decreasing the agent's sharing proportion (the piece-rate), the expected utility increases by the amount of

$$\left[\frac{\partial E\left[y \middle| Q(a^*(0),0)\right]}{\partial a} - C'(a^*(0))\right]\frac{da}{dr} + \left[\frac{\partial E\left[y \middle| Q(a^*(0),0)\right]}{\partial e} - \Psi'(0)\right]\frac{de}{dr}$$

The first bracket is zero by the definition of $a^*(0)$ and the second bracket is obviously positive. But, when $r = 1$, from (8), we obtain

$$\frac{de}{dr} = -\frac{\dfrac{\partial E\left[y \middle| Q(a^*(0),0)\right]}{\partial e}}{\Psi''(0)} < 0$$

Thus by marginally decreasing r, the principal can obtain more expected utility than she does with $(a^*(0), 0)$. So, this is a contradiction. The proof of $e^0(*) > 0$ can be shown in the same way.

Q.E.D.

Proposition 4 together with Proposition 3 leads to the optimality of a linear contract in a bilateral moral hazard setting. In a traditional agency model, where the principal is excluded from the production process, a linear contract that treats the agent as a residual claimant has been suggested as one of the first best contract. The principal who is not engaged in production can get the first best solution by charging the agent a fixed rent, combined with the piece -rate of 100 %. However, if the principal gets involved in the production process, by investing her own effort, such a contract does not work efficiently. When the principal also participates in the production process, she herself behaves as an agent after the wage contract is settled and in order to induce the agent to take the right action, she must convince him of her action strategy. Yet receiving a fixed rent cannot perform this job effectively, (at least in the short run,) resulting in the incentive loss on her side. Consequently (as a static framework) in such a team production between the principal and the agent, there is no way for the principal to establish full incentive without hurting her own commitment. Only the second best is obtainable and the simple sharing rules we frequently observe in small profit-sharing firms perform well.

The result of Proposition 4 depends crucially upon the assumption of $\psi'(0) = 0$. If we relax this assumption and assume that $\frac{\partial E\left[y \middle| Q(a^*(0),0)\right]}{\partial e} \le \Psi'(0)$, we will get the corner solution $r^* = 1$. This is because the principal cannot increase her expected utility by marginally decreasing the piece-rate r from 1, when the marginal cost of the principal *at zero effort* is sufficiently high. In this case, the optimal linear scheme is $r^* = 1$ and $s^* = \overline{U} + C(a^*(0)) - E\left[y \middle| Q(a^*(0), 0)\right]$, which induces $a^0(*) = a^*$ and $e^0(*) = 0$ at the second best.

Finally, specifying actual proportions to be shared between the principal and the agent depends on the production function, that is, on the degree to which each player marginally contributes to the outcome. The next section clarifies the argument and provides comparative statics with a simple example.

3. An example

The principal installs her effort e and the agent provides his effort a . Both efforts are assumed to be unobservable by the other party. The true amount (or the quality) of products is a function of both efforts, $Q(e,a)=\alpha\ln a+(1-\alpha)\ln e$, but the measured products is the true amount plus error term, that is, $y=Q(e,a)+\theta$ and $E(\theta)=0$, where E denotes the expectation operator. Since the price of products is normalized to one, $F\left(y\middle|Q(e,a)\right)$ denotes the revenue distribution with given effort choices of the principal and the agent. We specify the effort costs of the principal and the agent as follows; $\Psi(e)=\psi e$, and $C(a)=ca$.

The principal maximizes her expected profit denoted by

$$\alpha\ln a+(1-\alpha)\ln e-Ew(y)-\psi e \tag{14}$$

When the principal can enforce the worker the amount of effort, the optimal combination of inputs is $(a^*,e^*)=\left[\frac{\alpha}{c},\frac{1-\alpha}{\psi}\right]$ that generates the marginal benefit of each input equal to the marginal cost, and the principal pays ca^* as a lump sum payment to the agent.[14]

However, when the principal cannot force the amount of effort directly or indirectly, the principal should consider the incentive constraint of the agent. From Proposition 3, any combination of Nash equilibrium actions can be implemented by the linear sharing rule $w(y)=ry+s$. Thus, if $(a^0,e^0)\in C^0$, then (a^0,e^0) should satisfy

$$r\frac{\alpha}{a^0}=c \tag{15}$$

$$(1-r)\frac{1-\alpha}{e^0}=\Psi \tag{16}$$

In Figure 1, the principal's indifferent curves are drawn on (a,e) space and the set of $(a^0,e^0)\in C^0$ is drawn as a straight line, since $e^0=\frac{1-\alpha}{\psi}\left[1-\frac{ca^0}{\alpha}\right]$ from (15) and (16). Point F denotes the first best action combination $\left[\frac{\alpha}{c},\frac{1-\alpha}{\psi}\right]$ and Point E denotes the second best $\left(a^0(*),e^0(*)\right)$, which gives the maximum level of expected utility to the principal among $(a^0,e^0)\in C^0$. In order to find the second best action combination $\left(a^0(*),e^0(*)\right)$, it is enough to find the corresponding $r^0(*)$ since each action combination in the straight line corresponds to each value of $0\leq r\leq 1$ in $w(y)=ry+s$. For example, as r increases from zero to one, the action combination moves down along the line from A to B.

By plugging (15) and (16) into (14) and taking the derivatives with respect to r , we get the first order condition on the optimal sharing rate (piece-rate) $r^0(*)$

$$\frac{\alpha}{r^0(*)}-\frac{1-\alpha}{1-r^0(*)}-\alpha+(1-\alpha)=0 \tag{17}$$

Solving (17), we get

$$r^0(*)=\frac{\alpha-\sqrt{\alpha(1-\alpha)}}{2\alpha-1} \tag{18}$$

Now computing the marginal rate of substitution between a and e on the indifference curve of

[14] The individual rationality level of the agent is normalized to zero.

the principal, we get $MRS_{ae} = \dfrac{\alpha \frac{1}{a^0} - c}{(1-\alpha)\frac{1}{e^0} - \Psi}$. Similarly, we get $MRT_{ae} = \dfrac{1-\alpha}{\alpha}\dfrac{c}{\psi}$ as the marginal rate of transformation between a and e on C^0 .We can easily check that the marginal rate of substitution equals the marginal rate of transformation at $\left(a^0(*), e^0(*)\right)$.

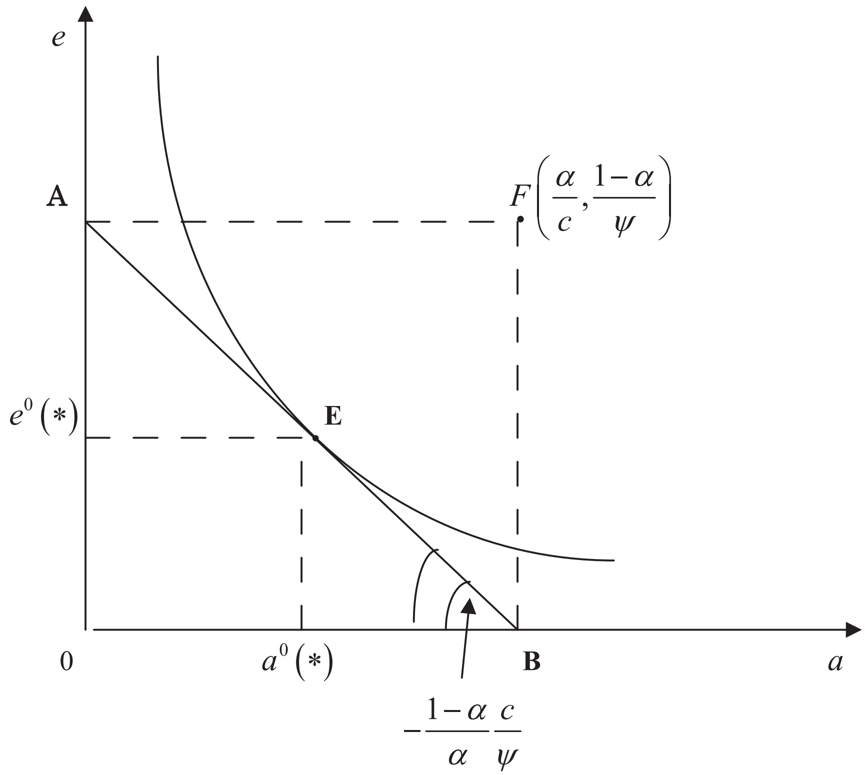

Figure1 an Example: The Second Best Solution

We can easily check $0 \le r^0(*) \le 1$ and from (18) we obtain

$$\frac{dr^0(*)}{d\alpha} = \frac{1 - 2\sqrt{\alpha(1-\alpha)}}{2\sqrt{\alpha(1-\alpha)}(2\alpha-1)^2} > 0 \tag{19}$$

since both the numerator and the denominator are positive for $0 < \alpha < 1$. (19) shows that as the relative proportion the worker's effort contributes to the production increases, his sharing proportion also increases. Some readers may think that an intuitive explanation for this result may be that it is fair that the worker receives more when he contributes more, but this intuition ignores the *incentive aspect* of the underlying wage contract. The right interpretation is that *the incentive weight* should move to the worker when he relatively contributes more to the output. To do so, the worker's sharing proportion (piece-rate) should increase. As a result, his effort provision increases in equilibrium.

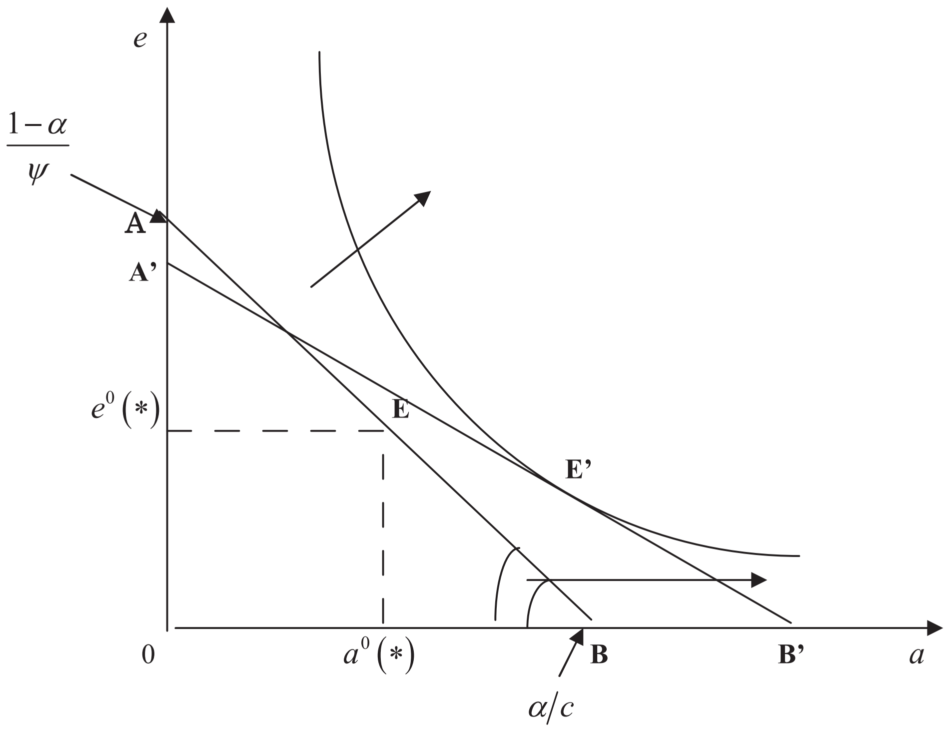

Figure2 The effect of increasing α on the optimal solution

If $\alpha = 1$, which represents the traditional principal-agent model, $r^0(*) = 1$ and the first best will be achieved. Thus, this model includes the traditional principal agent model as a special case. Finally, the fixed wage: $s^0(*)$ will be decided by satisfying the agent's individual rationality constraint, that is,

$$\int \left(r^0(*)\, y + s^0(*) \right) f\left(y \middle| Q\left(e^0(*), a^0(*) \right) \right) dy = c a^0(*) \tag{20}$$

4. Commitment Problem and its Solution under Repeated Relationships

4.1 A Repeated Contracting Model and Trigger Equilibrium

We have examined so far the bilateral moral hazard situation under the fully court-enforced contracting model. However, there exist many contracting elements that are "relational"- governed by the contracting parties rather than being fully reliant on the courts. Here, we shall introduce the idea of *relational contracts*[15]-contracts that are *self-enforced through repeated interaction* between the principal and the agent, rather than pure court-enforced contract. We can use repeated game methods to analyze the model.

If we investigate our problem in a repeated game framework, the principal may be able to

[15] For a theoretical model for relational contracting, see Levin (2003), which studies self-enforced incentive contracts under imperfect observability, using a model of repeated agency. His model does not analyze a bilateral moral hazard situation whereas our model does. Baker, Gibbons and Murphy (2002) examine an interaction between asset ownership (*who owns the asset*) and relational contracts both within (*employment*) and between firms (*outsourcing*). Our framework belongs to spot and relational employment relationship within firms (the principal owns the asset). Suzuki (2000) deals with a spot and relational outsourcing model between firms (i.e., Holdup and its solution).

solve the commitment problem under certain conditions. In the original model analyzed so far, the only observable (and verifiable) variable was the total output y and the principal could write the initial contract $w(y)$ at the beginning, and commits herself to the scheme, at least in one period. Now, we assume that though not observing the other parties' effort choices, each party observes the effort composition $q = Q(a,e)$, *before* it is buffered by a state of nature θ (measurement error or random shock) in each period. It is assumed to be *unverifiable*. Next, suppose that the principal and the agent set a target $\overline{q} = Q(a^*,e^*)$, where a^*, e^* are the first best effort choices defined in Section 2.2, which maximize the total surplus with no incentive constraints. Now, let us consider the following trigger type strategies. For the principal, in period $t = 1$ of the game, she offers a 100% piece-rate wage scheme $r^* = 1$, $s^* = E\left[y \middle| Q(a^*,e^*)\right] - C(a^*) - \overline{U}$ i.e. $w(y) = y - s^*$, and then she chooses the first best effort e^*. Afterwards, if the *history* in period t is $(r_\tau^*, s_\tau^*) = (1, s^*)$ and $q_\tau \geq \overline{q} = Q(a^*, e^*)$, $\forall \tau = 1, 2, \cdots, t-1$, then in period t, $(r_\tau^*, s_\tau^*) = (1, s^*)$ and she chooses the first best effort e^*. Otherwise, if there is either a cheating about the contract offer on the principal's side or the event that $q_t < \overline{q}$ occurred at least once, then she plays the static second best strategy $\left\{w^{0*}(y), e^0(*)\right\}$ satisfying (11) and (13) forever. Similarly for the agent, in period $t = 1$, if the principal offers a 100% piece-rate wage scheme, then he provides the first best effort a^*, characterized by $\dfrac{\partial E\left[y \middle| Q(a^*,e^*)\right]}{\partial a} = C'(a^*)$. Afterwards, if $(r_\tau^*, s_\tau^*) = (1, s^*)$, $\forall \tau = 2, 3, \ldots, t$ and $q_\tau \geq \overline{q} = Q(a^*,e^*)$, $\forall \tau = 1, 2, \ldots 1, t-1$, then he chooses the first best effort a^*. Otherwise, that is, if there is either a cheating about the contract offer on the principal's side or the event that $q_t < \overline{q}$ occurred at least once, then he plays the static Nash $a^0(*)$ forever.

These strategies are a variant of *trigger strategies*. We shall check that these strategies form a self-enforcing agreement (a subgame perfect Nash equilibrium: SPNE) in a game theoretic sense. The crucial point is that the principal must have an incentive not to deviate on the first best effort choice e^*, *after* setting the initial incentive contract $w(y) = y - s^*$. In other words, the principal's *commitment* about her optimal effort choice e^* must be *credible*.[16] Now, check her Deviation Incentive (DI) after offering $w(y) = y - s^*$ to the agent. Her static objective is to maximize $\int s^* f\left(y \middle| Q(a^*,e)\right) dy - \Psi(e)$. Thus, the *optimal* deviation is to choose $e^* = 0$, and the DI is $\left[s^* - \Psi(0)\right] - \left[s^* - \Psi(e^*)\right] = \Psi(e^*) - \Psi(0) = \Psi(e^*)$. While, by the fact that $q_t = Q(a^*, 0) < \overline{q} = Q(a^*,e^*)$, the strategies of both the principal and the agent would switch into the repetition of the static equilibrium $\left\{w^{0*}(y), \left(a^0(*), e^0(*)\right)\right\}$ from the next period on[17] and she would lose the following *continuation payoff*

$$\frac{1}{1-\delta}\Big[\left\{E\left[y \middle| Q(a^*,e^*)\right] - \Psi(e^*) - C(a^*) - \overline{U}\right\}$$
$$-\left\{E\left[y \middle| Q\left(a^0(*), e^0(*)\right)\right] - \Psi\left(e^0(*)\right) - C\left(a^0(*)\right) - \overline{U}\right\}\Big]$$

[16] The observation that under employment (integration) the identity of the party tempted to deviate is the principal is also consistent with the analysis by Baker, Gibbons and Murphy (2002). Of course, their model deals with transactional relationships, consisting of the upstream unit and the downstream unit, and derives neither the (static) optimal incentive schemes analyzed so far nor a characterization of the optimal solution (Proposition 5 below) when combined with relational contracts in the repeated relationships.

[17] Baker, Gibbons, and Murphy (1994) argue that there are important interactions between the relational contract and the formal contract, with the formal contract affecting the sustainability of the relational contract by defining the fallback positions of the parties. In their model, there are formal contracts ($w(y) = ry + s$ in this paper) and relational contracts (self-enforcing agreement, equilibrium trigger strategies between the principal and the agent), and the repetition of the static second best equilibrium forms the fallback position as "Nash threat" in the repeated game framework. In that sense, the structure of our model is close to theirs, but their model does not analyze the *bilateral moral hazard situation and ours does*.

$\equiv \frac{1}{1-\delta}\Delta TS$, where δ is the discount factor $\left(0 \leq \delta < 1\right)$.

Hence, the principal has an incentive to commit herself on her first best effort e^* if and only if

$$\Psi\left(e^*\right) \leq \frac{\delta}{1-\delta}\Delta TS \Leftrightarrow \delta \geq \frac{\Psi\left(e^*\right)}{\Psi\left(e^*\right)+\Delta TS}$$

Then, since the principal's commitment to choosing e^* is credible, the agent trusts her and has an incentive to choose the first best effort a^*, characterized by $\frac{\partial E\left[y \middle| Q\left(a^*, e^*\right)\right]}{\partial a} = C'\left(a^*\right)$.

Last, it is obvious that the principal does not have an incentive to deviate from the contract offer $\left(r^* = 1, s^*\right)$, given the *history* that the cooperative equilibrium $\left(a^*, e^*\right) \gtrsim \left(a^0\left(*\right), e^0\left(*\right)\right)$ has been played so far, and given the *expectation* that if she did not cheat this time, the cooperative equilibrium would be preserved in the future.

4.2 The role of *reputation or trustworthiness*

We shall interpret the above story from the viewpoint of the role of reputation. Concern with one's *reputation* may be an effective check on the principal's ex post opportunism, where she deviates from e^* to 0 after promising the first best effort e^* when offering the initial contract $w\left(y\right)$, and the reputational concern may overcome the temptations to renege. It can indeed achieve the same result as actual commitment, as above for a certain range of the discount factor δ. In a world of costly and incomplete contracting, *trust* is crucial to realizing valuable transactions. Thus, the concern with getting a bad reputation that takes away future possibilities for profitable transactions can limit reneging. Effectively, it removes the incentives for opportunistic behavior by creating a cost $\frac{\delta}{1-\delta}\Delta TS$ offsetting the short-term gain of opportunistic behavior $\Psi\left(e^*\right)$.

As Milgrom and Roberts (1992) pointed out, the value of a good reputation increases with the number of times it may be used, and in the employment relationship with a stable, long-lived firm, the firm will have a longer horizon than the employee (worker). In our model, we can view the principal as the infinitely-lived firm, and the agent as the short-lived employee. Thus, the principal, who put in a position of power (e.g. offering a short-term contract in each period), will have a stronger incentive to build and maintain a reputation with the agent for using power (the right of offering the initial contract) well. Concretely, if the agent expects that the principal is trustworthy and provides the first best effort as promised, he will also have an incentive to provide the first best effort under $w\left(y\right) = y - s^*$, where $y = Y\left(Q\left(a, e^*\right), \theta\right)$, *given that the principal provides the first best effort* e^*.

4.3 The full solution of the problem, including the second best cases for

$$\Psi\left(e^*\right) > \frac{\delta}{1-\delta}\Delta TS \Leftrightarrow \delta < \frac{\Psi\left(e^*\right)}{\Psi\left(e^*\right)+\Delta TS} \text{ evaluated at } \left(a^*, e^*\right) \text{ (the first best levels)}$$

If the discount factor δ is low, the first best optimum may not satisfy the principal's self-enforcing constraint. In such a case the maximum effort level e^{**} that satisfies the principal's self-enforcing constraint will be supported in equilibrium as *the constrained optimum*.

Now, define $\overline{\delta}$ by $\overline{\delta} = \dfrac{\Psi(e^*)}{\Psi(e^*) + \Delta TS}$. For $0 \leq \delta < \overline{\delta}$, the principal's self enforcing constraint is binding and her first best effort level cannot be elicited in equilibrium. The intuitive reason is that when the effort level e^* promised by the principal becomes too large, the agent expects the principal to cheat on the effort provision, because it pays the principal to simply suffer a loss in reputation and save on the effort cost. In such a case, it may be valuable to use a formal contract (enforceable piece-rate) r, where $0 < r < 1$, to implement the second best effort e^{**}.

To show this, for the discount factor $\delta \in [0, \overline{\delta})$, suppose that the principal offers a piece-rate scheme $w(y) = ry + s$, where $0 \leq r \leq 1$ and $s = \overline{U} + C(a) - rE[y|Q(a,e)]$, and then she chooses an effort level e^{**}. Now, check whether the principal's *commitment* about her effort choice e^{**} is *credible*. First, the principal's static objective is to maximize

$$\int((1-r)y + s) f(y|Q(a,e))dy - \Psi(e) = (1-r)\int yf(y|Q(a,e))dy + s - \Psi(e)$$

Thus, the *optimal* deviation is to choose e' given the agent's equilibrium effort choice a^{**}, and the *Deviation Incentive* (DI) is

$$(1-r)\left\{E[y|Q(a^{**},e')] - E[y|Q(a^{**},e^{**})]\right\} - \left\{\Psi(e') - \Psi(e^{**})\right\}$$

Note that when $r = 1$, it is $\Psi(e^{**}) - \Psi(0) = \Psi(e^{**})$, since the optimal deviation is to choose $e' = 0$. Next, the *Continuation Loss* is,

$$\frac{1}{1-\delta}\Big[\left\{E[y|Q(a^{**},e^{**})] - \Psi(e^{**}) - C(a^{**}) - \overline{U}\right\}$$
$$-\left\{E[y|Q(a^0(*),e^0(*))] - \Psi(e^0(*)) - C(a^0(*)) - \overline{U}\right\}\Big]$$
$$\equiv \frac{1}{1-\delta}\left[E\pi_p(e^{**},a^{**};r^{**},s^{**}) - E\pi_p(a^0(*),e^0(*);r^0(*),s^0(*))\right]$$ [18]

which gives the present discounted value of the continuation payoff when they keep the cooperative equilibrium path minus the continuation payoff attained by the repetition of the static (second best) equilibrium under the optimal formal contract $(r^0(*), s^0(*))$ in the absence of a relational contract. Hence, the self-enforcement constraint is,

$$(1-r)\left\{E[y|Q(a^{**},e')] - E[y|Q(a^{**},e^{**})]\right\} - \left\{\Psi(e') - \Psi(e^{**})\right\}$$
$$\leq \frac{\delta}{1-\delta}\left[E\pi_p(e^{**},a^{**};r^{**},s^{**}) - E\pi_p(a^0(*),e^0(*);r^0(*),s^0(*))\right]$$

where the left-hand side of the inequality is *DI or Reneging Temptation,* while the right-hand side is *Continuation Loss*.

Now, suppose that the following equality holds at e^{**},

$$(1-r)\left\{E[y|Q(a^{**},e')] - E[y|Q(a^{**},e^{**})] - \Psi(e') - \Psi(e^{**})\right\}$$
$$= \frac{\delta}{1-\delta}\left[E\pi_p(e^{**},a^{**};r^{**},s^{**}) - E\pi_p(a^0(*),e^0(*);r^0(*),s^0(*))\right]$$

and that $e^{**} = \max\{e | \text{Deviation Incentive} = \text{Continuation Loss}\}$. Then, we notice that the DI (reneging temptation) is decreased, by *explicitly using the "formal contract" part r , that is, not* $r = 1$ but $0 < r < 1$, and the level of e^{**} that satisfies the above equality is increased. We also notice that while the level of e^{**} increases, a^{**} determined by the first order condition

[18] Refer to an example in Section 3 for the notations and $(a^0(*), e^0(*))$ and $(r^0(*), s^0(*))$

$r\frac{\partial E\left[y\middle|Q\left(a^{**},e^{**}\right)\right]}{\partial a} = C'\left(a^{**}\right)$ is nothing but *second best*, because $r < 1$.

Totally differentiating the above equality which equalizes DI with "Continuation Loss", and putting it in order by using the *envelope theorem*, we obtain the relation.

$$\frac{\partial e^{**}}{\partial \delta} = \frac{\text{Continuation Loss} + \text{Deviation Incentive}}{(1-\delta)\times \text{Deviation Incentive}} > 0 .$$

We also see that the total surplus function $TS := E\left[y\middle|Q(a,e)\right] - \Psi(e) - C(a) - \overline{U}$ has *Single Crossing Property (SCP)* in (a,e), because $TS(a,e)$ has the property that $\frac{\partial^2 TS(a,e)}{\partial a \partial e} = \frac{\partial^2 E\left[y\middle|Q(a,e)\right]}{\partial a \partial e} > 0$. The intuition says that higher e increases the expected benefit of raising a. Then, based on Topkis's Theorem and the 'monotone comparative statics' technique, supposing that $a^*(e) = \arg\max_a TS(a,e)$, the maximizer $a^*(e)$ is *strictly increasing* in e.[19] These observations immediately yield the following proposition.

<u>Proposition5:</u>

The *optimal* effort levels induced in equilibrium is classified into two categories, depending on δ.

(1) If $\overline{\delta} \le \delta < 1$, then the <u>first best</u> effort levels (a^*, e^*), defined in Section 2.1, are attained in a self-enforcing way.

(2) If $0 \le \delta < \overline{\delta}$, then the principal's effort level is

$$e^{**} = \max\left\{e \middle| \text{Deviation Incentive} = \text{Continuation Loss}\right\}$$

and the agent's effort level is the <u>second best</u> $a^{**}(e^{**})$, given the principal's effort level e^{**}.

From the observations so far, we immediately have the following corollary.

<u>Corollary</u>

(1) There exists a $\overline{\delta} \in (0,1)$ such that for $\delta \in \left[\overline{\delta}, 1\right)$, the <u>first best</u> effort levels (a^*, e^*) are induced.

(2) For $0 \le \delta < \overline{\delta}$, the principal's effort level e^{**} and the agent's effort level $a^{**}(e^{**})$ are *strictly increasing* in δ, if two activities a and e are sufficiently *complementary*[20]. Moreover, a^{**} and e^{**} are both second best, and implemented by a combination of *formal* $(0 < r^{**} < 1)$ and *relational* contracts.

4.4 A remark on other equilibria

As is well known in the Folk Theorem literature (e.g. Fudenberg and Maskin (1986)) and the repeated contracting literature (e.g. Macloed and Malcomson (1989)), the above 'cooperative' equilibrium is not the only possible equilibrium in the repeated game. As an example, an outcome where the principal offers the static second best contract in every period, and then the

[19] This *strict* monotone comparative statics is due to the fact that the derivative $\frac{\partial E\left[y\middle|Q(a,e)\right]}{\partial a}$ exists and strictly increasing in e. See Edlin and Shannon (1998). The original result in monotone comparative statics by Topkis (1998) says that when the objective function satisfies *Supermodularity in* (a,e), which is equivalent to $\frac{\partial^2 E\left[y\middle|Q(a,e)\right]}{\partial a \partial e} \ge 0$, *maximizers are nondecreasing in the parameter value, in this case*,

[20] Milgrom-Roberts (1992) Chapter4 characterizes the complementarity relationship among activities in mathematical terms, and presents its application in organization design.

principal and the agent play the static Nash equilibrium $\left(a^0(*), e^0(*)\right)$ in every period is also a subgame perfect Nash equilibrium. This equilibrium is just a repetition of the second best equilibrium of our original static model. The following self-fulfilling expectations will make a deviation by the principal from this equilibrium unprofitable. When the principal newly offers $w(y) = y - s^*$, the agent believes that even if he chooses the first best effort a^*, characterized by $\frac{\partial E\left[y \middle| Q(a^*, e^*)\right]}{\partial a} = C'(a^*)$, the principal will not choose the first best effort e^*, characterized by $\frac{\partial E\left[y \middle| Q(a^*, e^*)\right]}{\partial e} = \Psi'(e^*)$, but the static best response $e^* = 0$. The principal also believes that even if she chooses the first best effort e^*, the agent will not choose the first best effort a^*, given e^*, but rather another effort level, characterized by $\frac{\partial E\left[y \middle| Q(a, 0)\right]}{\partial a} = C'(a)$. She also believes that, even if she provides the first best effort e^* despite the agent's expectation, he will neither change the belief nor provide the first best effort a^*, given e^* in subsequent periods. Then, it would be optimal for the principal not to deviate from the above 'non-cooperative' equilibrium

5. Extensions

5.1 Identification of Deviations and Contingent Control Shift

In our model, it is assumed that the actions (e, a) chosen by the principal and the agent affect the distribution of the outputs only through the one-dimensional factor $q = Q(e, a)$. The characterization of the static optimal contract heavily follows from such decomposition. Can we extend the model to the case that a and e can influence the distribution of outputs *in different ways*, as in the formulation of $y = Y(a, e, \theta)$ and $F\left(y \middle| a, e\right)$? A guess is that the first best may be implemented when the assumption of the one-dimensional decomposition is dropped. This is because the parties may be able to utilize the different effects of their actions on the outputs for designing the wage scheme to induce them to choose the appropriate actions. Being able to identify the agent who has not deviated may be enough to deter deviation. Legros-Matthews (1993) assume *deterministic output* like $q = Q(a, e)$, and so perfect identification of shirking. Their idea relies on the ability of a third party to always *identify* the player who did not shirk, and to undertake big transfers in her favor from the shirker, whenever somebody shirked.

We combine this idea with Aghion-Bolton (1992)'s "contingent control shift" idea. It is a financial contracting model with one investor and one entrepreneur, and the distribution of profit depends on the *verifiable* realization of the state of nature and on an *unverifiable* action to be taken. Initial contracts are to specify "divisions of final profit", like $\left\{ry + s, (1 - r)y - s\right\}$, and "contingent control allocations", which put somebody in charge, in other words, specify who can take the action, contingent on the state of nature. Aghion-Bolton (1992)'s *contingent-control contract* gives all the *financial return* to the investor, and *control* to the investor in one state and to the entrepreneur in another state. Here, in our model, we assume the existence of a referee, such as non-strategic and neutral external auditors,[21] who knows who has (or has not) deviated and shifts the authority of offering initial formal contract $w(y)$ to the agent, when the

[21] Thus, another enforcement problem still remains: the referee may behave strategically.

principal has deviated and the agent has not. In this setting, the principal's continuation payoff after reneging today is the present discounted value of her reservation utility $\frac{\delta}{1-\delta}\overline{U}_P$.

Therefore, comparing the principal's continuation losses after reneging in the two regimes, it is greater under "contingent control shift" la Aghion-Bolton (1992) by

$$\frac{\delta}{1-\delta}\left[\left\{E\left[y\middle|Q\left(a^*,e^*\right)\right]-\Psi\left(e^*\right)-C\left(a^*\right)-\overline{U}_A\right\}-\overline{U}_P\right]-$$
$$\frac{\delta}{1-\delta}\left[\left\{E\left[y\middle|Q\left(a^*,e^*\right)\right]-\Psi\left(e^*\right)-C\left(a^*\right)-\overline{U}_A\right\}-\left\{E\left[y\middle|Q\left(a^0(*),e^0(*)\right)\right]\right.\right.$$
$$\left.\left.-\Psi\left(e^0(*)\right)-C\left(a^0(*)\right)-\overline{U}_A\right\}\right]$$
$$=\frac{\delta}{1-\delta}\left\{E\left[y\middle|Q\left(a^0(*),e^0(*)\right)\right]-\Psi\left(e^0(*)\right)-C\left(a^0(*)\right)-\overline{U}_A-\overline{U}_P\right\}$$
$$=\frac{\delta}{1-\delta}\left\{\text{Static Second Best Surplus}\right\}.$$

This will tend to satisfy the principal's self-enforcing (promise keeping) constraint today, under the "contingent control shift" regime, and thus the first best will be attained for a broader range of discount factors.

5.2. Renegotiation-Proofness Problem and Corporate Restructuring

As in most repeated game models, an infinite number of equilibria exist in our model. Among them, we have focused on the trigger equilibrium that induces the efficient outcome. This trigger equilibrium, as explained in the Section 4, requires *punishment*, if especially the principal cheats, in the form of a trigger being pulled that the agent does not trust her anymore. The punishment (i.e., reversion into the repetition of static equilibrium, i.e., the second best equilibrium in our static framework), however, is itself inefficient. Then the principal could always claim to have made a mistake, and offer to the agent another efficient contract requiring punishment in the future if cheating (deviation) occurs. If the agent believes that, when cheating occurs, the punishment strategy will be renegotiated as this time, then the original contract cannot be self-enforcing.[22] Thus, it is important that the agent (employee) accepts a first best contract $w^*(y)=y-s^*$ followed by the first best actions (e^*,a^*) enforced by the trigger equilibrium only from the principals that have *never cheated*.

In summary, the problem of renegotiation proofness implies that the existence of a self-enforcing contract requires a convention that, once the principal loses her reputation, it is very hard to reacquire. We could say that Japanese firms in the post war period started from scratch in terms of their employer-employee relations and thus were able to begin with efficient self-enforcing agreements. Further, through the high growth economy period, there had existed "High-Reputation=Good Performance by good labor relations" firms and "Low-Reputation = Bad Performance by bad labor relations" firms, more exactly, an infinite number of forms of

[22] For the renegotiation problem in repeated games, see e.g., Farrell and Maskin (1989) and Van Damme (1989). They show that although all symmetric equilibrium payoffs can be supported by strategies which have the bang-bang property, i.e., strategies that use only the highest and lowest perfect equilibrium payoffs as continuation payoffs, these strategies cannot be (weakly) renegotiation proof, since the highest continuation payoff strictly pareto dominates the lowest one, and then the original efficient path itself cannot be supported in equilibrium due to the *incredibility of punishment*. Then, the only (weakly) renegotiation proof equilibrium is to play Static Nash equilibrium all the time, where there is only one continuation payoff.

labor relations between the principal and the agent.[23]

Through the 1990's until recently, we have often observed the fact that an increasing number of Japanese firms are losing their competitive advantage due to downturns in the relations between the employer and the employee, which are difficult to recoup. Concretely, in many firms, the principal (manager, or more generally, board of directors) are losing her trust, due to inefficient management since the "bubble" economy. Nonetheless, as the renegotiation-proofness shows, once a firm falls into the inefficient equilibrium, it is very difficult to recover from the situation through the direction by the same or similar top management. Hence, in such cases, instead of "self-enforcing" recovery, a completely new way must be introduced. In other words, they must stop the original game and restart "a New Game". From this point of view, we could understand the corporate reform ("Nissan Revival Plan") of Nissan Automobile Company by Carlos Ghosn, instead of the Japanese management.[24]

5.2.1. Another interpretation of δ

As for a parameter δ that is dealt as a discount factor in the model, we can adopt another interpretation of a *cost to renegotiation of cooperation* in a multi-period game. When δ is low, even if each player deviates from the cooparative equilibrium path, both parties can renegotiate *with low cost* to restart the cooparative equilibrium again, which largely weakens the punishment against deviation. Expecting this future renegotiation, neither party would not try to cooperate any more today. This interpretation seems to explain the old Nissan labor practice, where 'renegotiation' occurred very often and there did not exist any 'commitment' on the side of management. In that sense, Carlos Ghosn could be understood as a 'commitment device' to cooperative behavior on the principal's side, more simply speaking, a new principal with high δ . Given this, the players would not anticipate the same fate that the first best equilibrium will be collapsed and reputation will be ruined again, because the commitment by the principal with high δ is very credible.

6. Concluding Remarks

This paper investigated the characteristics of a wage contract and the impossibility of the first best, when the principal and the agent, after the wage contract is settled, constitute a team and produce the output together by investing their own efforts that are not observable with each other. In a traditional principal-agent model, where the principal does not play the role of another productive agent, the only issue is how the incentive can be induced from the agent. However, in a bilateral moral hazard framework like team production, another issue is added, that is, how the principal can convince the agent of her own effort incentive. Both incentives are trade-off and only the second best is possible even with a risk-neutral agent (in the static framework). It is shown that a simple linear contract we observe frequently in the real world can implement such second best efficiency.

The actual sharing proportion of each party crucially depends on the characteristics of production function, especially, the ratio that each party contributes to the production. A simple example shows that, as the agent's relative contribution to production gets larger, his shar-

[23] As for the theoretical literature on this viewpoint, see e.g., Okuno-Fujiwara (1989).

[24] Kikusawa (2006) suggests that the manager who failed in the past could not make the right decision in the present and would fail again, and so should be replaced, by using a mental accounting approach in Economic Psychology.

ing proportion increases to impose more weight on his effort incentive.[25]

Then, in Section 4, we introduced the idea of *relational contracts*-contracts that are *self-enforced through repeated interaction* between the principal and the agent, rather than pure court-enforced (formal) contract, and showed that in a repeated game version of the model, the commitment problem mentioned above could be solved, and a first best outcome could be achieved through both parties taking *trigger strategies* depending on a public (but unverifiable) signal. Then, we give an interpretation in the viewpoint of the 'reputation' mechanism and a qualitative characterization about the optimal solution induced in equilibrium for all possible ranges of discount factor. It can be viewed as analyzing how the implementable incentives change under the connection with the *formal contract* and the *relational contract*, for all discount factors. Then, we interpreted the model in a context of 'multiple equilibria', and derived an implication for recent corporate restructuring problem from the viewpoint of 'Renegotiation-proofness' and 'Commitment'.[26]

Acknowledgements

I would like to thank the participants at *International Conference on Game Theory and Applications*, China, August 2002, *Contract Theory Workshop East* at Hitotsubashi University 2003, Micro Workshop at University of Tokyo 2003, *Econometric Society European Meeting*, Stockholm, Sweden, August 2003, and two referees of this journal for their useful comments. Financial support for this research was provided through Grant-in-Aid for Scientific Research by the Japan Society for the Promotion of Science (No. 17530192). Finally, I would like to thank Hosei Institute of Comparative Economic Studies for the constant support through the project entitled "A Comparative Analysis of Governance: A Game Theoretic and Contract Theoretic Approach".

REFERENCES

Aghion, P and P. Bolton (1992) "An Incomplete Contracts Approach to Financial Contracting". *Review of Economic Studies*. 59. 473 - 494

Baker, G., R.Gibbons, and K. Murphy. (1994) "Subjective Performance Measures in Optimal Incentive Contracts" *Quarterly Journal of Economics*, November 1125-1156.

Baker, G., R.Gibbons., and K. Murphy. (2002) "Relational Contracts and the Theory of the Firm." *Quarterly Journal of Economics*, February 39-84

Bhattacharyya, S and Lafontaine, F (1995) "Double-sided Moral Hazard and the Nature of Share Contracts," *Rand Journal of Economics* 26(4), 1995, 761-781.

Bolton, P and M. Dewatripont (2005) *Contract Theory* MIT Press

Bull,C. (1987) "The existence of Self-Enforcing Implicit Contracts," *Quarterly Journal of Economics,* 102, 147-159.

Chaterjee, K and W.Samuelson, (1983). "Bargaining under Incomplete Information", *Operations Research*. 31:835-51.

[25] By using this example in Section 3, we could exercise a comparative statics to the repeated game analysis in Section 4. For example, as α (the relative contribution by the agent to the total product) increases, i.e., $1-\alpha$ decreases, the 'deviation incentive' by the principal will decrease, and the 'continuation loss', which is equivalent to the difference of the total surpluses that the principal can extract through the initial contract, will also increase. Hence, 'self-enforcement constraint' by the principal will be greatly *relaxed*, and so we will have the result on a critical discount factor $\overline{\delta}$ such that $\partial\overline{\delta}/\partial\alpha < 0$.

[26] For a related research, see Suzuki (2005).

Edlin, A and Shannon, C (1998) “Strict Monotonicity in Comparative Statics,” *Journal of Economic Theory*, 81, July, 201-219.

Farrell, J., and E.Maskin (1989) “Renegotiation in Repeated Games,” *Games and Economic Behavior* 1, 327-360.

Fudenberg, D and E.Maskin (1986) “The Folk Theorem in Repeated Games with Discounting or with Incomplete Information,’ *Econometrica*, 54,533-554.

Holmstrom, B. (1982) “Moral Hazard in Teams.” *Bell Journal of Economics*, 13,324-340.

Kikusawa,K (2006) “Mental Accounting: Science in Leader’s Behavior” *Diamond Harvard Business Review*, February, 94-108.

Legros,P and S.Matthews (1993) “Efficient and Nearly Efficient Partnerships,” *Review of Economic Studies* 60, 3, 599-611

Levin, J. (2003) “Relational Incentive Contracts,” *American Economic Review*, 93(3), June, 835-847.

Macleod, B. and J. Malcomson. (1989) “Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment,” *Econometrica*, 57, 447-480.

Milgrom, P and J.Roberts. (1992) *Economics, Organization and Management*, Prentice-Hall, Englewood Cliffs.

Nandeibam, S (2002), “Sharing Rules in Teams”, *Journal of Economic Theory*, 107, 407-420.

Okuno-Fujiwara, M (1989) “On Labor Incentives and Work Norm in Japanese Firms” *Journal of the Japanese and International Economies* 3, 367-384

Rayo, L (2001) “Relational Contracting and Team Incentives” mimeo-graphed, Stanford University.

Suzuki, Y. (2000) “Thinking of ‘Corporate Governance in Japanese Firms’ in Game Theory”, in M. Nakayama, S. Muto, and Y. Funaki, eds., *Thinking in Game Theory*, Yuhikaku, in Japanese.

Suzuki, Y. (2005) “Commitment, Renegotiation, and Relational Contracts in Procurement Relationships”, Mimeo-graphed, Hosei University. in Japanese.

Topkis, D. (1998) *Supermodularity and Complementarity* Princeton University Press.Princeton.

Van Damme, E (1989) “Renegotiation-Proof equilibrium in Repeated Prisoners’ Dilemma,” *Journal of Economic Theory* 47,202-217.