Review of Economic Studies (1994) **61**, 109–129 0034-6527/94/00070109$02.00

Renegotiation and Optimality in Agency Contracts

CHING-TO ALBERT MA
Boston University

First version received January 1991; *final version accepted August* 1993 (*Eds.*)

We analyse renegotiation in a hidden action principal-agent model. Contract renegotiation offers are made by the agent. A refinement is imposed on the principal's beliefs: if precisely one action is optimal with respect to both the principal's and the agent's contracts, the principal believes that that action has been taken. With the refinement imposed, perfect-Bayesian equilibrium allocations are identical to the second best in the classical principal-agent model without renegotiation. When renegotiation is led by the agent and when equilibria satisfy the refinement, equilibrium allocations are ex ante efficient.

1. INTRODUCTION

Incentive contracting is a common mechanism for promoting ex ante investments and for avoiding ex post opportunistic behaviour in a long-term or continuing economic relationship. The classical principal-agent model (Holmstrom (1979), Grossman and Hart (1983)) is a natural framework for this class of problem. In such a model, a risk-averse agent performs an action on behalf of a risk-neutral principal; the agent's action is unobservable and affects the principal's revenue in a random fashion. The derivation and characterization of efficient compensation schemes have been the major focus in the principal-agent literature. Generally, the incentives to motivate a costly action from the agent take the form of "profit sharing:" the agent's reward is made contingent on his random performance outcomes.

Recently, economists have questioned the classical model's implicit assumption that contracting parties can commit to a compensation scheme or a contract. The point is the following. Because the agent's action is unobservable and has uncertain consequences, the provision of incentives by a compensation scheme imposes risks on the agent. But contracting parties may realize that incentives are unnecessary after some point, and may attempt to repeal the inefficient components of a contract in time before its final execution. For example, when they know that the agent's action has already been taken but that its outcome is still uncertain, they may renegotiate the contract. A rational agent, however, anticipates such renegotiation at the outset. Therefore, his action decision will be based on the expected revised contract. Thus, a primary consideration is the possibility of contract renegotiation.

In a recent paper, Fudenberg and Tirole (1990) model renegotiation as the principal's lack of commitment to a contract. In their paper, a contract consists of a menu of compensation schemes, one for each possible action. The principal may propose an alternative contract after the agent has completed his unobservable action but before the action's uncertain consequence is realized; this new contract may replace the earlier one if both agree. After renegotiation but still before the action's result is known, the agent must pick a scheme from the agreed contract.

Fudenberg and Tirole show that in equilibrium either the agent chooses the least-cost action or he chooses an action according to the realization of a probability distribution on the set of actions. In other words, when the principal can make renegotiation offers, the set of implementable actions becomes smaller: a costly action with probability one is not implementable.[1]

This result can be explained as follows. If in equilibrium the agent has performed an action with probability one, then at the renegotiation stage the principal must offer a full insurance compensation scheme. Hence, when choosing among his actions, the agent expects that he will be fully insured. This implies that if in equilibrium the agent takes an action with probability one, it must be the least-cost action. A non-degenerate probability distribution on actions may be implementable because the principal need not make a full insurance renegotiation offer in that situation. If the agent picks an action according to some distribution, actions in the distribution support then determine the agent's possible preferences towards compensation schemes—his "types"—at renegotiation. When uncertain about the agent's actual action and hence his preference, at renegotiation the principal can offer a menu of compensation schemes with dissimilar risk characteristics to screen the different types of the agent. Therefore, incentives for costly actions remain.

Fudenberg and Tirole also show that for some agent utility functions, a contract offering a higher rent to the agent expands the set of implementable action distributions. It is then possible that in equilibrium the agent earns rents above his reservation utility. In summary, Fudenberg and Tirole's results stand in contrast to those in models without renegotiation: in the classical principal-agent model the principal need never consider random actions and the agent never obtains more than his reservation utility when his utility function is separable with respect to income and effort.

In this paper, we investigate renegotiation in the principal-agent model when the agent, rather than the principal, proposes the renegotiation contract; otherwise, our model is essentially similar to the Fudenberg-Tirole model. In particular, a contract in this paper is also defined as a menu of compensation schemes. We find that under a belief restriction the equilibrium allocation is identical to that in the classical principal-agent model without renegotiation. When renegotiation offers are made by the agent, not only can costly actions arise in equilibrium, but equilibria satisfying a refinement must be second best.

The pair of models in Fudenberg and Tirole (1990) and here illustrate the lack of robustness in renegotiation procedures: different extensive forms of renegotiation in the principal-agent model yield entirely different equilibrium allocations. On the positive side, however, our result indicates that the incentive-efficient allocation of the classical principal-agent model is still the contracting outcome when the agent makes renegotiation offers and when equilibria satisfy a belief refinement. In situations where the principal can credibly refrain from making contract revision proposals, the standard principal-agent model is applicable regardless of renegotiation.

What explains the second-best result when the agent, instead of the principal, makes renegotiation offers? Observe that in the Fudenberg-Tirole analysis, at renegotiation the principal's belief about the agent's action must remain identical to her initial belief. By contrast, in our model, at renegotiation the principal's belief can depend on the agent's new contract offer. For example, suppose the principal initially proposes an incentive contract. Then the principal can believe that the agent has taken an inferior action if she

1. Ma (1991) considers multi-period contracts. In that model renegotiation occurs after an output has been observed in the first period, The paper shows that implementing a costly action with probability one is still feasible, but due to binding renegotiation-proof constraints the second best cannot be achieved.

receives a full insurance new proposal from the agent, and she may indeed reject it. Anticipating a rejection of his renegotiation offer, the agent chooses a costly action.

Although this argument might also imply a multiplicity of equilibrium allocations in our model, there actually exists a second-best unique equilibrium allocation if the principal's belief satisfies the following restriction: when the principal's initial contract and the agent's renegotiation contract support the same, unique best action for the agent, then the principal must believe that the agent has performed this action. Thus the principal believes that certain weakly dominated strategies are never used by the agent: the agent proposing a contract that supports the same, unique best action as the principal's initial contract, but not taking this action is weakly dominated.

Two other recent papers have studied contract renegotiation in the principal-agent framework. Hermalin and Katz (1991) consider renegotiation after some signal about the agent's action becomes known to the principal and the agent; the information from this signal, however, is non-verifiable, and hence the initial or renegotiation compensation schemes cannot be based on it. Most strikingly, when this signal takes the form of the agent's actual action, Hermalin and Katz prove that the first best may be achieved through renegotiation whether the principal or the agent makes the renegotiation offer. More generally, they show that renegotiation can be used to exploit the non-verifiable information from this signal to improve the welfare of the contracting parties. In our model, the principal receives no new signal about the agent's action at the renegotiation stage apart from the inference she draws from the agent's offer.

Matthews (1993) studies the same problem as in this paper, except that a contract in his model consists of a single compensation scheme, rather than a menu, so equilibria in this paper are excluded in his analysis. Nevertheless, the second best is achieved if equilibria satisfy a belief restriction. In Matthews's construction of the equilibrium, the principal's initial contract sells the entire production process to the agent—a sales contract—at a price that gives the principal her second-best payoff. In equilibrium the agent offers the second-best compensation scheme at renegotiation, which is then accepted by the principal.

In the next two sections we lay out the classical principal-agent model and the renegotiation model. The analysis is in Section 4. The bulk of the paper focuses on the case where the agent can choose between two actions. We then discuss how the result can be generalized when the agent has an arbitrary, finite number of actions. Conclusions are drawn in Section 5.

2. THE PRINCIPAL-AGENT MODEL

In this section we describe a standard contracting model. The principal has a stochastic production process which requires the agent's effort or action. This action is unobservable to the principal. Initially, we assume that the agent can choose between two effort levels, e_1 and e_2, with $e_1 < e_2$. This assumption simplifies presentation; we later discuss extensions to the case of an arbitrary, finite number of effort levels. Each action determines a distinct probability distribution on a vector of n possible revenues $(R_1, \ldots, R_h, \ldots, R_n) \equiv R$. The vector $(\pi_1(e_i), \ldots, \pi_h(e_i), \ldots, \pi_n(e_i)) \equiv \Pi(e_i)$ denotes the probability distribution induced on R by effort e_i, $i=1, 2$.

The principal is risk neutral, and has a payoff equal to the realized revenue less any transfer to the agent. The agent's utility function is additively separable with respect to money and effort; he is risk averse with respect to money. If the agent receives a monetary income I from the principal and expends effort e, his utility is $U(I) - G(e)$, where U is strictly increasing and strictly concave, and G is strictly increasing. Finally, the agent's reservation utility from not participating is normalized at zero.

If the agent's effort were observable, the principal would bear all the risks. However, if it is not, the agent will provide the least effort when all risks are absorbed by the principal. To motivate the agent to perform a costly action, appropriate incentives must be provided through a compensation scheme.

The earlier principal-agent literature has studied contracting outcomes when contract renegotiation is ruled out; see Holmstrom (1979) and Grossman and Hart (1983). In these papers, an incentive scheme specifies a transfer v_h from the principal to the agent contingent on revenue R_h. A scheme $v\equiv(v_1, \dots, v_h, \dots, v_n)$ is said to *implement* action e_i if given the scheme, action e_i maximizes the agent's expected utility and gives him at least the reservation utility. An *incentive-efficient scheme* for action e_i is a scheme that implements e_i at minimal cost to the principal.

Formally, the incentive-efficient scheme for e_i, $v^{i*}=(v_h^{i*})$, solves Programme $P(e_i)$: Choose v_h, $h=1, \dots, n$, to

$$\text{minimize} \sum \pi_h(e_i)v_h$$

subject to

$$\sum \pi_h(e_i)U(v_h)-G(e_i)\geqq\sum \pi_h(e_j)U(v_h)-G(e_j), \qquad i\neq j, \tag{1}$$

$$\sum \pi_h(e_i)U(v_h)-G(e_i)\geqq 0, \tag{2}$$

where the summation operator here and throughout the paper applies to index $h=1, \dots, n$. Programme $P(e_i)$ and subsequent programmes in the paper are assumed to possess solutions. The objective function in $P(e_i)$ is the principal's expected cost from the contract. Constraint (1) is the incentive constraint; it ensures that e_i is a best action. Constraint (2) guarantees the agent his reservation utility. Because the agent's attitude towards risk is independent of his action due to the separable utility function, random compensations are never optimal. Therefore v_h is assumed non-random. (Formal proofs that stochastic v_h's are suboptimal can be found in Holmstrom (1979) and Arnott and Stiglitz (1988).)

We discuss some properties of v^{i*}. By the strict concavity of U, the incentive-efficient scheme v^{i*} must be unique.[2] The reservation utility (2) must bind. If the incentive constraint (1) does not bind, then the incentive-efficient scheme for e_i achieves first-best risk sharing. Clearly (1) does not bind for e_1, since it is the least-cost action. Hence v^{1*} consists of a constant wage. The incentive constraint (1) often binds for the incentive-efficient scheme for e_2; Proposition 3 in Grossman and Hart (1983) gives conditions for this.[3] However, our results do not rely on incentive constraints being binding; even the full insurance property of v^{1*} is not used for the proofs.

The principal's expected utilities from implementing actions with incentive-efficient schemes are

$$\sum \pi_h(e_i)[R_h-v_h^{i*}]. \tag{3}$$

An *allocation* is defined as a probability distribution on the agent's actions and compensation schemes. The seond-best allocation in the classical principal-agent model is the action

2. The uniqueness property only applies to transfers that are paid with strictly positive probability under action e_i. Transfers that never occur under e_i can be specified arbitrarily, provided the reservation and incentive constraints hold. The nature of this uniqueness property applies to all solutions of optimization programmes in this paper. The uniqueness of incentive-efficient schemes is most easily derived by the Grossman-Hart method of transforming the minimization instrument from dollars to utilities: with the transformation, the programme becomes the minimization of a strictly convex function subject to a set of linear constraints.

3. Roughly, these conditions say that (1) binds if $\Pi(e_1)$ and $\Pi(e_2)$ have identical supports.

and its incentive-efficient scheme that maximizes the principal's expected utilities (3). We assume that the principal prefers implementing e_2 to e_1:

$$\sum \pi_h(e_1)[R_h - v_h^{1*}] < \sum \pi_h(e_2)[R_h - v_h^{2*}]. \tag{4}$$

Then the expression on the right hand side of (4) is the principal's second-best payoff. We call $[e_2; v^{2*}]$ *the second-best allocation.* With (4) and absent contract renegotiation, the principal proposes incentive scheme v^{2*}, the agent performs action e_2 and is compensated according to v^{2*}. For future use we let

$$\alpha = \sum \pi_h(e_1)[R_h - v_h^{1*}] \quad \text{and} \quad \beta = \sum \pi_h(e_2)[R_h - v_h^{2*}].$$

For the renegotiation model, we need to extend the standard approach to include random actions and their implementation. In this paper, a *contract* is defined as a menu containing an arbitrary, finite number of compensation schemes. To implement random actions, the principal initially proposes a contract, then the agent picks an action according to a probability distribution on his set of available actions. Next, the agent selects one of the schemes from the contract. Finally, the principal's revenue is realized, and the agent is compensated according to the scheme he has selected from the contract.

To study incentive-efficient contracts for random actions, first note that the number of schemes in the contract can be taken as the number of actions available to the agent. Hence, consider contract $\{v^i, i=1, 2\}$, where $v^i = (v_h^i)$ and v_h^i is the agent's compensation if he picks scheme i and the revenue becomes R_h. We say that an agent is type e_i when he has chosen action e_i. Without loss of generality, v^i is a most preferred scheme in $\{v^1, v^2\}$ for a type e_i agent, $i=1, 2$. A contract $\{v^i\}$ *implements* action e_2 with probability p, $0<p<1$, if given $\{v^i\}$, choosing actions e_1 and e_2 with probabilities $1-p$ and p is optimal[4] and gives the agent at least his reservation utility. An *incentive-efficient contract* for action e_2 with probability p is a contract that implements e_2 with probability p at minimal cost to the principal. So an incentive-efficient contract for e_2 with probability p, $0<p<1$, is a solution to Programme $P(e_2, p)$: Choose v_h^i, $h=1, \ldots, n$, $i=1, 2$, to

$$\text{minimize} \sum (1-p)\pi_h(e_1)v_h^1 + \sum p\pi_h(e_2)v_h^2$$

subject to

$$\sum \pi_h(e_1)U(v_h^1) - G(e_1) = \sum \pi_h(e_2)U(v_h^2) - G(e_2) \geqq 0, \tag{5}$$

$$\sum \pi_h(e_i)U(v_h^i) \geqq \sum \pi_h(e_i)U(v_h^j), \qquad i, j=1, 2, i \neq j. \tag{6}$$

The objective function is the principal's expected cost under the contract $\{v^1, v^2\}$ when the agent picks action e_2 with probability p and a type e_i agent selects v^i from the contract. Constraint (6) says that a type e_i agent will not gain by selecting scheme v^j, $i \neq j$. Constraint (5) ensures that the agent is indifferent between actions e_1 and e_2, and obtains his reservations utility.

Proposition 0. *For any p, $0 \leqq p \leqq 1$, and incentive-efficient contract for action e_2 with probability p is $\{v^{1*}, v^{2*}\}$, where v^{i*} is an incentive-efficient scheme for action e_i, $i=1, 2$.*

For completeness, a proof of this proposition is provided in the Appendix.[5] The incentive-efficient implementation of random actions is achieved simply by pairing together

4. The agent must be indifferent between e_1 and e_2.
5. Lemma 2.1 in Fudenberg and Tirole (1988) stated the same result.

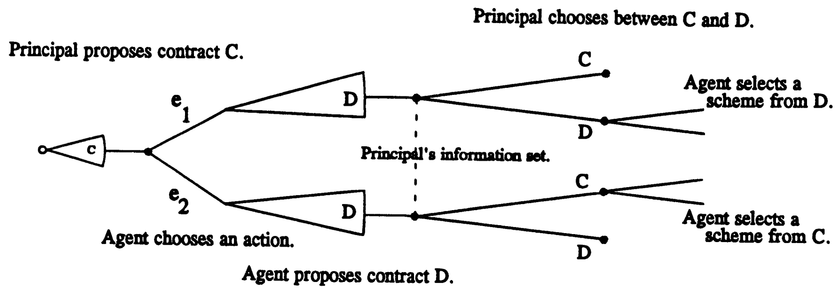

FIGURE 1: Game I

the incentive-efficient scheme for actions in the support of the distribution. Although Programme $P(e_2, p)$ for incentive-efficient contracts of random actions has considered only non-degenerate distributions by restricting p strictly between 0 and 1, Proposition 0 holds for the two limits of $p=0$ and $p=1$ as well: both e_1 and e_2 are optimal actions given contract $\{v^{1*}, v^{2*}\}$, and v^{i*} is incentive-efficient for e_i. Also, Proposition 0 extends to the case of an arbitrary, finite number of actions: an incentive-efficient contract for a distribution on a finite number of actions consists of incentive-efficient schemes for actions in the distribution support.

Since the incentive-efficient contract for random actions is independent of the agent's probability distribution on actions, the expected cost of implementing e_2 with probability p is a weighted average of the cost of implementing action e_1 and action e_2 with weights $1-p$ and p respectively. Thus, the principal's payoff is monotonic in p. In the standard analysis, the principal need never consider implementing random actions, and generically will strictly prefer a non-random action.

3. A RENEGOTIATION MODEL

We now describe the contract renegotiation game. Initially the principal proposes a contract C. If the agent accepts C, he chooses an unobservable action from the set $\{e_1, e_2\}$. Also, the agent may propose another contract D. Without knowing the agent's action, the principal then chooses between contracts C and D. Next, the agent picks a compensation scheme from the contract chosen by the principal. A revenue then becomes realized and the transfer is executed according to the scheme selected by the agent. In this game, renegotiation is conducted under imperfect information (since the principal never observes the agent's action), and is led by the party with superior information (the agent).

Two extensive forms can be defined for the renegotiation game. These extensive forms, Games I and II, are illustrated[6] in Figures 1 and 2. Obviously they have the same normal form; the only difference between them is the timimg of the agent's action and his contract counter-proposal. In Game I the agent's contract renegotiation offer is made after he has chosen an action; in Game II the agent announces a contract and then picks an action.[7] Notice that in both games the principal observes the agent's contract counteroffer but not his action, and the agent chooses an action

6. For brevity, both the agent's option of not participating and the execution of transfers after the agent has selected the final scheme have been left out in the figures.

7. There is a third extensive form in which the agent picks an action and a contract simultaneously, but this is equivalent to Game I.

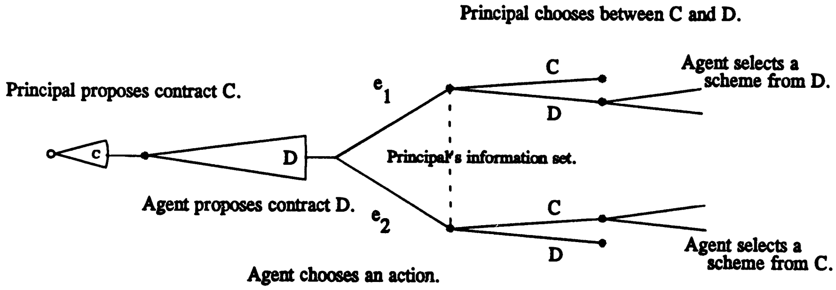

FIGURE 2: Game II

before the principal selects the final contract. One can identify subgames defined by the principal's contract C in both games; these are the only subgames in Game I. But in Game II one can identify another class of subgames—those defined by the pair ofljycontracts (C, D) announced by the two parties.

It is well known that perfect-Bayesian equilibria (see Fudenberg and Tirole (1991) for a definition) are not robust to "irrelevant" transformations of the game tree; see for example Kohlberg and Mertens (1986). The sets of perfect-Bayesian equilibria in Games I and II need not be identical. In Game II, because of the existence of subgames defined by (C, D), perfect-Bayesian equilibria are identical to subgame-perfect equilibria and require that the agent's action and the principal's contract choice be mutual best responses for *any* given pair of contracts (C, D). However, perfect-Bayesian equilibria in Game I do not require this. For an arbitrary C, any perfect-Bayesian equilibrium in subgame C in Game II is a perfect-Bayesian equilibrium in subgame C in Game I, but the converse is not true.[8]

We use Game I for two reasons. In the next section we will show that there is a unique equilibrium allocation when perfect-Bayesian equilibria in Game I satisfy a belief restriction. This restriction is implied by perfect-Bayesian equilibria in Game II. Because any perfect-Bayesian equilibrium in Game II is a perfect-Bayesian equilibrium in Game I, the (unique) allocation in equilibria satisfying the refinement in Game I must be the unique perfect-Bayesian equilibrium allocation in Game II. Thus, with the refinement imposed, our results are independent of the choice of the particular extensive form. Second, Game I is more in line with the Fudenberg-Tirole (1990) model where the renegotiation contract is proposed by the principal after an action has been chosen.

At the renegotiation stage, our model fits exactly an "informed-principal" model by Maskin and Tirole (1992). But our entire game is actually richer because of the stages *before* renegotiation. This leads to two differences between the informed-principal game and ours. First, the informed-principal model is one of incomplete information—the party who possesses private information at the start of the game (the principal in their model) proposes contracts. Our model is one of imperfect information—the agent learns his private information (his choice of an unobservable action) during the play of the game. Different efficiency concepts should be applied to games with different information structures. This point has been made by Holmstrom and Myerson (1983) (especially p. 1807).

8. Consider a given equilibrium in Game II; this equilibrium specifies the principal's and the agent's moves for all subgames (C, D). To construct a perfect-Bayesian equilibrium in Game I, one simply uses the same equilibrium strategies for the perfect-Bayesian equilibrium, and defines the principal's beliefs to be the agent's equilibrium actions for all subgames (C, D) in Game II. Perfect-Bayesian equilibria in Game I need not be subgame-perfect equilibria in Game II; examples are some of the equilibria in Proposition 1.

Results in the informed-principal model are related to weakly interim and interim efficiencies,[9] the relevant efficiency notions in games of incomplete information. In our model, the appropriate concept is *ex ante efficiency*, which corresponds to our definitions of incentive-efficient contracts for actions and the second-best allocation. An interim efficient allocation is generally ex ante inefficient. So, even if equilibria in the informed-principal model lead to interim efficient allocations, ex ante efficiency in our model does not immediately follow. Second, in our model the uninformed player's belief is generated endogenously. With this richer structure our analysis can be based on ex ante considerations: at the renegotiation (interim) stage of our game, players' beliefs can still depend on the (ex ante) properties of the status quo contracts. This is infeasible in the informed-principal model, since ex ante preferences are not defined for each player and the status quo contract is exogenous.

4. ANALYSIS

A contract in the renegotiation game consists of a finite menu of compensation schemes from which the agent can choose. The principal's strategy is an initial contract offer C, and a rule that specifies the probability of accepting the agent's contract D. Given any initial contract C and the agent's proposal D, the principal's belief structure is a probability distribution on the set of actions. The agent's strategy consists of a choice of action and a contract proposal as a function of the principal's initial contract, and a choice of compensation scheme from any contract. Given a perfect-Bayesian equilibrium, an equilibrium allocation is defined by the agent's equilibrium actions(s), and the agent's equilibrium reward scheme(s).

In the literature, the notion of "renegotiation-proof" contracts has been used. Although we do not need to appeal to it directly, and in fact we need to consider contracts that are not renegotiation-proof, the next proposition can be related to the concept of "weakly renegotiation-proof contracts", due to Maskin and Tirole (1992, p. 24). To rephrase their concept here, we say that a contract C is *weakly renegotiation-proof* if there exists an equilibrium in subgame C in which either the agent proposes C again or his counteroffer is rejected. Our first result characterizes equilibria when the principal proposes the contract $C^0 \equiv \{v^{1*}, v^{2*}\}$, and shows that C^0 is weakly renegotiation-proof and that the second-best allocation $[e_2; v^{2*}]$ is an equilibrium allocation.

Proposition 1. *Consider the subgame defined by the principal's initial contract* $C^0 \equiv \{v^{1*}, v^{2*}\}$. *For any* p, $0 \leqq p \leqq 1$, *there is a perfect-Bayesian equilibrium in subgame* C^0 *in which the agent proposes* C^0, *performs actions* e_1 *and* e_2 *with probabilities* $(1-p)$ *and* p *respectively, and is compensated according to* v^{i*} *when he picks action* e_i, $i = 1, 2$. *That is, contract* C^0 *is weakly renegotiation-proof.*

Proof. First, from Proposition 0 and inequality (6), it is optimal for a type e_i agent to pick v^{i*} from the contract $C^0 \equiv \{v^{1*}, v^{2*}\}$. Second, consider any contract renegotiation offer from the agent, $D = \{w^1, w^2, \ldots, w^K\}$, where $w^i = (w^i_h)$ pays the agent w^i_h contingent on revenue R_h. Define $k_D(i) \equiv \min[\operatorname{argmax}_k \sum \pi_h(e_i) U(w^k_h), k = 1, \ldots, K]$; that is $k_D(i)$ is the smallest index of a type e_i agent's most preferred schemes in D. We now construct an

9. A contract is weakly interim efficient if (i) it induces truth-telling for the agent, and (ii) there exists no other Pareto-dominating (among all types of agent) contract satisfying (i) that the principal prefers regardless of her belief. The definition of interim efficiency is well known; see Holmstrom and Myerson (1983).

equilibrium in which the agent proposes $D=C^0$, performs action e_2 with probability p, and selects v^{i*} from C^0 if he has performed e_i.

We start by describing the agent's equilibrium strategy. In subgame C^0, he takes action e_2 with probability p and offers $D=C^0$. When choosing among schemes in C^0, he picks scheme v^{i*} if he has taken action e_i. If he is to choose a scheme from any other contract $D=\{w^1, w^2, \ldots, w^K\}$, he picks $w^{k_D(i)}$ if he has taken action e_i, $i=1,2$.

Next, we describe the principal's strategy in subgame $C^0 \equiv \{v^{1*}, v^{2*}\}$. We begin by describing the principal's belief on the agent's action for any renegotiation offer $D=\{w^1, w^2, \ldots, w^K\}$. First, compute the set of best actions under D. That is, for any contract D, define the set

$$\Psi(D)=\{e_i : i \in \operatorname{argmax}_j \sum \pi_h(e_j)U(w_h^{k_D(j)})-G(e_j)\}.$$

If $\Psi(D)=\{e_i\}$, the principal believes that the agent has selected action e_i, $i=1,2$. If $\Psi(D)=\{e_1, e_2\}$, the principal believes that the agent has chosen e_2 with probability p. In words, if given D the agent has a unique best action, the principal believes that this action has been taken. If given D the agent is indifferent between e_1 and e_2, the principal believes that action e_2 has been taken with probability p, where $0 \leqq p \leqq 1$.

Given this system of beliefs and the agent's strategy, the principal chooses between contracts C^0 and D to maximize her utility. If $\Psi(D)=\{e_i\}$, then select C^0 (or reject D) if and only if

$$\sum \pi_h(e_i)v_h^{i*} \leqq \sum \pi_h(e_i)w_h^{k_D(i)}. \tag{7}$$

If $\Psi(D)=\{e_1, e_2\}$, then select C^0 (or reject D) if and only if

$$(1-p)\,\pi_h(e_1)v_h^{1*}+p\pi_h(e_2)v_h^{2*} \leqq (1-p)\pi_h(e_1)w_h^{k_D(1)}+p\pi_h(e_2)w_h^{k_D(2)}. \tag{8}$$

We now verify that these strategies and beliefs form a perfect-Bayesian equilibrium in subgame C^0. Clearly the principal's strategy is optimal given her beliefs and the agent's strategy in subgame C^0. The agent's selection rule from any contract is also optimal. If the agent offers $D=C^0$, performing e_2 with probability p is optimal, since given C^0 he is indifferent between both actions and attains his reservation utility.

It remains to show that the agent cannot gain by proposing $D \neq C^0$. These are two possibilities: (a) e_i is the unique best action given D, and (b) both actions are optimal given D. Consider case (a). Since the agent can always obtain his reservation utility by proposing C^0, we can assume that his optimal action given D is worth at least the reservation utility. By assumption e_i is the unique best action, we have

$$\sum \pi_h(e_i)U(w_h^{k_D(i)})-G(e_i) \geqq \sum \pi_h(e_j)U(w_h^{k_D(j)})-G(e_j).$$

Also, a type e_j agent prefers scheme $w^{k_D(j)}$ to scheme $w^{k_D(i)}$. Hence

$$\sum \pi_h(e_j)U(w_h^{k_D(j)}) \geqq \sum \pi_h(e_j)U(w_h^{k_D(i)}).$$

These two inequalities imply that $w^{k_D(i)}$ satisfies the incentive constraint (1) and the reservation utility constraint (2) in programme $P(e_i)$. Hence (7) must hold; the principal will reject D. It follows that the agent cannot gain by proposing such a contract. For case (b) we can use the same argument as in case (a) to conclude that $\{w^{k_D(1)}, w^{k_D(2)}\}$ satisfies all the constraints in programme $P(e_2, p)$. Therefore (8) must hold. Again, the principal will reject D. In sum, the agent cannot gain by offering $D \neq C^0$. ||

The intuition for Proposition 1 is this. C^0 is an incentive-efficient contract for any action when a type e_i agent chooses v^{i*} from it. When she receives a renegotiation offer, the principal believes that the agent has taken an optimal action relative to the new contract. She rejects it if her utility from it is less than that from C^0. But this means the principal will only accept a renegotiation contract when it is incentive-efficient and Pareto superior to C^0. Hence if the principal accepted the renegotiation offer, the agent had to receive less than his expected utility from C^0.

Proposition 1 shows that any probability distribution on actions can be part of a perfect-Bayesain equilibrium in subgame C^0. The second-best allocation $[e_2;\ v^{2*}]$ is an equilibrium allocation (when $p=1$). Existence of a perfect-Bayesian equilibrium in the entire Game I is also established. The principal's equilibrium expected utility is bounded by her expected utility in the second-best allocation. For $p=1$, the continuation equilibrium in subgame C^0 gives her this expected utility. Thus, given this continuation equilibrium, it is a best response for the principal to propose C^0 in Game I. It follows that Game I possesses a perfect-Bayesian equilibrium.

Proposition 1 contrasts with the Fudenberg-Tirole (1990) result that equilibrium action must be random when the principal makes renegotiation offers. Although the agent's expected utilities are identical in all the equilibria in Proposition 1, the principal strictly prefers the second-best equilibrium. The contribution of this paper is to show that with a belief restriction the second best is the unique equilibrium allocation in the *whole* game.

Before continuing, we introduce a belief restriction:

> *A Refinement.* Suppose the principal initially proposes contract C, and later the agent proposes an alternative contract D. Suppose the agent has an identical unique best action given C or D. Then the principal must believe with probability one that the agent has taken this action.

The interpretation of this belief restriction is as follows. Suppose the agent responds to the principal's contract C by proposing contract D. For each of these contracts, action e_i and a choice of the most preferred scheme from the contract are the only optimal moves for the agent.[10] What should the principal believe? Notice that e_i is the agent's best action *independent* of the principal's choice of contracts. The refinement insists that the principal must believe that e_i has been taken. Formally it says the principal does not believe that the agent uses certain weakly dominated strategies. Proposing contract D against C and taking an action other than e_i is a weakly dominated strategy in the renegotiation game.[11] One can also adopt a "forward induction" interpretation for this refinement. If in subgame C the agent had intended to perform $e_j \neq e_i$, he would not have proposed a contract D where e_j was sub-optimal under either C or D. Thus, by proposing D, he signals that his action must be e_i.

Notice that the refinement does not refer to any equilibrium; it is based on a dominance comparison. The refinement is only applicable to a subset of all contract pairs. It places no restriction on the principal's belief either when there are multiple best actions given a contract, or when best actions under the principal's and the agent's contracts are not identical. Equilibria described in Proposition 1 satisfy the refinement vacuously, since under C^0 the agent is indifferent between his actions.

The refinement is implied by perfect-Bayesian (or subgame-perfect) equilibria in Game II. Consider a subgame in Game II defined by contracts C and D. Our refinement then translates to the requirement that when there is a unique action that strictly dominates

10. That is, from notation in the proof of Proposition 1, we have $\Psi(C)=\Psi(D)=\{e_i\}$.

11. It is a strictly dominated strategy in the subgame defined by C.

other actions in subgame (C, D), the principal believes that the agent will take such an action in subgame (C, D). Clearly this must be true in a Nash equilibrium in subgame (C, D).

Our goal is to show that equilibria satisfying the refinement must lead to a second-best allocation. We first show that in the subgame defined by an appropriate perturbation of C^0, the principal's payoff in any perfect-Bayesian equilibrium satisfying the refinement is smaller than, but arbitrarily close to, her second-best payoff.

We use a contract that provides strict incentives for the agent to perform e_2, $C^\varepsilon = \{v^{1*}, x^\varepsilon\}$, indexed by $\varepsilon > 0$, where $x^\varepsilon = (x_h^\varepsilon)$ solves the following programme: Choose $a_h, h = 1, \ldots, n$, to

$$\text{minimize} \sum \pi_h(e_2) a_h$$

subject to

$$\sum \pi_h(e_2) U(a_h) - G(e_2) \geqq \sum \pi_h(e_1) U(a_h) - G(e_1) + 2\varepsilon, \tag{9}$$

$$\sum \pi_h(e_2) U(a_h) - G(e_2) \geqq \varepsilon. \tag{10}$$

At a solution, the reservation utility constraint (10) must bind. From this, (9), (6), and $\sum \pi_h(e_i) U(v_h^{i*}) = G(e_i)$ we know that for contract C^ε a type e_1 [resp. type e_2] agent strictly prefers scheme v^{1*} [resp. x^ε]:

$$\sum \pi_h(e_1) U(v_h^{1*}) = G(e_1) > G(e_1) - \varepsilon \geqq \sum \pi_h(e_1) U(x_h^\varepsilon),$$

$$\sum \pi_h(e_2) U(x_h^\varepsilon) = G(e_2) + \varepsilon > \sum \pi_h(e_2) U(v_h^{2*}) \geqq \sum \pi_h(e_2) U(v_h^{1*}).$$

In contract C^ε the agent obtains utility ε [resp. zero] by performing e_2 [resp. e_1]. As ε tends to zero, x^ε tends to v^{2*}, and contract C^ε tends to contract C^0 in Proposition 1. We now analyse equilibria in subgame C^ε.

Proposition 2. *In any perfect-Bayesian equilibrium in subgame C^ε, the agent must choose e_2 with positive probability.*

Proof. The agent's equilibrium expected utility in subgame C^ε must be at least ε, because he can always offer C^ε again and perform e_2. Suppose the proposition is false; that is, suppose that in a perfect-Bayesian equilibrium in subgame C^ε the agent chooses action e_2 with zero probability. First, if in equilibrium the principal always picks C^ε, then from action e_1 the agent gets $\sum \pi_h(e_1) U(v_h^{1*}) - G(e_1) = 0 < \varepsilon$. This is a contradiction. Hence, in equilibrium the agent must offer a contract $D \neq C^\varepsilon$ with positive probability and the principal must accept it with positive probability.

Recall that given e_1 the principal's payoff from C^ε must be α, since selecting v^{1*} from C^ε is a type e_1 agent's unique best move. Therefore given e_1 her payoff from accepting the renegotiation contract D must be at least α. In equilibrium the agent selects among some schemes in contract D. At least one of these equilibrium schemes in D gives the principal at least α; otherwise she would not accept D. So, without loss of generality, let w^1 in contract D be chosen by the agent with positive probability, and given action e_1 scheme w^1 let the principal obtain at least α. Furthermore the agent's equilibrium expected utility from w^1 must be at least ε. These arguments imply that e_1 cannot be a best action given D: the agent's expected utility from w^1 and e_1 is

$$\sum \pi_h(e_1) U(w_h^1) - G(e_1) \geqq \varepsilon > 0,$$

which says w^1 strictly satisfies the reservation utility constraint (2) in programme $P(e_1)$. Recall that α is the principal's payoff from the solution to Programme $P(e_1)$. So when the principal obtains at least α from scheme w^1 and action e_1, scheme w^1 must violate the incentive constraint (1) in Programme $P(e_1)$. Hence e_1 cannot be an optimal action given D. But e_2 is the unique best action given contract C^ε. So independent of the principal's choice of contract, e_1 is sub-optimal. We conclude that in an equilibrium in subgame C^ε the agent cannot take e_1 with positive probability. This is a contradiction. ||

Define $\delta(\varepsilon)$ by $\delta(\varepsilon)=\sum \pi_h(e_2)(R_h-x_h^\varepsilon)$. Thus, $\delta(\varepsilon)$ is the principal's payoff under contract C^ε when the agent takes action e_2. Recall β is the principal's second-best payoff. From the definition of x^ε, we have $\delta(\varepsilon)<\beta$ for $\varepsilon>0$, and $\lim_{\varepsilon\to 0}\delta(\varepsilon)=\beta$.

What is the agent's maximum equilibrium expected utility in subgame C^ε? Which contract will achieve this? If the agent's offer is always rejected, then his equilibrium expected utility is ε. But the agent may improve his welfare by offering a contract acceptable to the principal. From Proposition 2 we know that action e_2 must be an optimal action given the equilibrium contract. Furthermore, if the principal believes with probability ϕ_i that the agent type's is e_i, $i=1, 2$, then her "reservation payoff" is the expected utility she can obtain from contract C^ε, $\phi_1\alpha+\phi_2\delta(\varepsilon)$. The agent's alternative contract must offer the principal at least this reservation utility.

The above discussion suggests that we should look for schemes that maximize the agent's expected utility subject to the constraints that e_2 is a best action, and that the principal obtains at least her reservation utility. Since we only know that e_2 must be an optimal action given the equilibrium contract, we must consider separately equilibria where e_2 is always performed, and where e_2 is chosen with probability p, $0<p<1$.

In equilibrium, the agent makes some contract counter-proposal. It is convenient to consider two possibilities: either this counter-proposal is always accepted by the principal, or it is accepted by the principal with positive probability less than one. Let us consider the first case. Suppose that e_2 is always performed, consider Programme $P(\varepsilon)$: Choose a_h, $h=1,\ldots,n$, to

$$\text{maximize} \sum \pi_h(e_2)U(a_h)-G(e_2)$$

subject to

$$\sum \pi_h(e_2)U(a_h)-G(e_2)\geqq \sum \pi_h(e_1)U(a_h)-G(e_1), \tag{11}$$

$$\sum \pi_h(e_2)a_h \leqq \sum \pi_h(e_2)x_h^\varepsilon. \tag{12}$$

Constraint (11) says that e_2 is a best action, and constraint (12) says the principal weakly prefers scheme (a_h) to x^ε given e_2. By the strict concavity of U, the solution to $P(\varepsilon)$ is unique. At a solution constraint (12) binds. Let $y^\varepsilon\equiv(y_h^\varepsilon)$ be the solution. Denote the optimized value of the objective function of Programme $P(\varepsilon)$ by $\gamma(\varepsilon)$. Hence, $\gamma(\varepsilon)=\sum \pi_h(e_2)U(y_h^\varepsilon)-G(e_2)\geqq\varepsilon$.[12] Observe that $\gamma(\varepsilon)$ is an upper bound on the agent's expected utility in equilibria with e_2 always performed.

Now we turn to equilibria in which e_2 is taken with positive probability $p<1$. Clearly, in these equilibria the agent must be indifferent between his actions, and the principal must weakly prefer accepting the agent's offer. So consider the following: Choose a_h^i, $i=1, 2$, $h=1,\ldots,n$, to

$$\text{maximize} \sum \pi_h(e_2)U(a_h^2)-G(e_2)$$

12. It is clear that $\gamma(\varepsilon)>\varepsilon$ and $y^\varepsilon\neq x^\varepsilon$ if and only if constraint (9) binds at a solution to that programme.

subject to

$$\sum \pi_h(e_1)U(a_h^1)-G(e_1)=\sum \pi_h(e_2)U(a_h^2)-G(e_2), \tag{13}$$

$$\sum \pi_h(e_i)U(a_h^i)\geqq\sum \pi_h(e_i)U(a_h^j), \qquad i,j=1,2,\ i\neq j, \tag{14}$$

$$\sum (1-p)\ \pi_h(e_1)a_h^1+\sum p\pi_h(e_2)a_h^2\leqq\sum (1-p)\ \pi_h(e_1)v_h^{1*}+\sum p\pi_h(e_2)x_h^{\varepsilon}. \tag{15}$$

Constraint (14) says that a type e_i agent selects scheme a^i; constraint (13) ensures that the agent is indifferent between actions. Finally, constraint (15) states that the principal weakly prefers $\{a^1, a^2\}$ to $\{v^1, x^{\varepsilon}\}$. Again, (15) must bind at a solution; the proof of Proposition 0 in the Appendix can be modified to verify this. The strict concavity of U implies that the solution, $\{z^1, z^2\}$, is unique. Denote by $\kappa(p, e)$ the value of the objective function at $\{z^1, z^2\}$; that is, $\kappa(p, \varepsilon)\equiv\sum \pi_h(e_2)U(z_h^2)-G(e_2)$ is an upper bound on the agent's expected utility if in equilibrium e_2 is performed with probability p.

Clearly, if $\kappa(p, \varepsilon)<\varepsilon$, then choosing e_2 with probability p cannot be part of an equilibrium since the agent's equilibrium expected utility in C^{ε} is at least ε. So if there is an equilibrium in which e_2 is chosen with probability $p<1$, then we must have $\kappa(p, \varepsilon)\geqq\varepsilon$.

Lemma 1. *Suppose $p<1$ and $\kappa(p, \varepsilon)\geqq\varepsilon$, then $\kappa(p, \varepsilon)<\gamma(\varepsilon)$.*

Proof. By hypothesis $\kappa(p, \varepsilon)\geqq\varepsilon$ and the agent is indifferent between the actions, so we know that $\sum \pi_h(e_1)U(z_h^1)-G(e_1)\geqq\varepsilon>0$. Moreover, from (13) and (14), we have $\sum \pi_h(e_1)U(z_h^1)-G(e_1)\geqq\sum \pi_h(e_2)U(z_h^1)-G(e_2)$. So scheme z^1 satisfies all constraints in the programme for the incentive-efficient contract for action e_1, $P(e_1)$. However, z^1 makes the reservation utility constraint (2) in $P(e_1)$ slack. Hence given e_1 scheme z^1 gives the principal less than α:

$$\sum \pi_h(e_1)z_h^1>\sum \pi_h(e_1)v_h^{1*}.$$

Since (15) must hold as an equality at the solution $\{z^1, z^2\}$, it follows that

$$\sum \pi_h(e_2)z_h^2<\sum \pi_h(e_2)x_h^{\varepsilon}.$$

From this inequality, and the fact that (13) and (14) imply

$$\sum \pi_h(e_2)U(z_h^2)-G(e_2)\geqq\sum \pi_h(e_1)U(z_h^2)-G(e_1),$$

we know that z^2 satisfies all constraints in programme $P(\varepsilon)$. However, z^2 makes constraint (12) slack, and therefore cannot be a solution to that programme. It follows that $\kappa(p, \varepsilon)\equiv\sum \pi_h(e_2)U(z_h^2)-G(e_2)<\sum \pi_h(e_2)U(y_h^{\varepsilon})-G(e_2)\equiv\gamma(\varepsilon)$. ‖

Intuitively, for each action we find the incentive-efficient scheme when the principal's status quo contract is $C^{\varepsilon}=\{v^{1*}, x^{\varepsilon}\}$. The best incentive scheme for the agent is the one for action e_2 with probability one. Contract $\{v^{1*}, x^{\varepsilon}\}$ would have been an incentive-efficient contract for all actions, including random ones, if ε were zero. With a strictly positive ε, it is not incentive-efficient except for action e_1. But the slacks in the contract $\{v^{1*}, x^{\varepsilon}\}$ are at the scheme for action e_2. The agent benefits most if these slacks are reduced and e_2 is always performed. If he randomizes between actions, the principal's expected cost of C^{ε} becomes a weighted average of costs from an incentive-efficient scheme and an incentive-inefficient one. The scheme v^{1*} in C^{ε} then limits the amount of slack that can be reduced. In sum, the Lemma says that in equilibria in which the agent's contract is always accepted, an upper bound for the agent's expected utility is $\gamma(\varepsilon)$. Moreover, since y^{ε} is a unique solution to $P(\varepsilon)$, the agent achieves $\gamma(\varepsilon)$ only if he is compensated according to scheme y^ε.

Let us turn to the second case; that is, equilibria in which the agent's proposal is accepted with positive probability less than one. Again, consider first those equilibria in which the agent performs e_2 with probability one. Can the agent achieve an expected utility higher than $\gamma(\varepsilon)$? Observe that y^ε is an incentive-efficient scheme for this action: it maximizes the agent's expected utility subject to the agent's incentive and the principal's reservation utility constraints.

Suppose the agent proposes a contract $D=\{w^1,\dots,w^K\}$, and without loss of generality assume that a type e_i agent picks scheme $(w_h^{k_D(i)})$ from D (where $k_D(i)$ is defined in the proof of Proposition 1). Suppose the principal chooses between C^ε and D with probability q and $1-q$. Then she must be indifferent between x^ε and $w^{k_D(2)}$ so we have

$$\sum \pi_h(e_2)[qx_h^\varepsilon+(1-q)w_h^{k_D(2)}]\leqq\sum \pi_h(e_2)x_h^\varepsilon,$$

which says that the principal's reservation constraint (12) is satisfied by the stochastic scheme that pays the agent according to x^ε and $w^{k_D(2)}$ with probabilities q and $1-q$ respectively.

From the agent's point of view, the principal's randomization over C^ε and D results in a stochastic compensation scheme: with probability q he will be compensated by x^ε, and with probability $1-q$ by $w^{k_D(2)}$. Nevertheless, this stochastic scheme must satisfy his incentive constraint (11) in Programme $P(\varepsilon)$:

$$\begin{aligned}\sum \pi_h(e_2)[qU(x_h^\varepsilon)&+(1-q)U(w_h^{k_D(2)})]-G(e_2)\\ &\geqq\sum \pi_h(e_1)[qU(v_h^{1*})+(1-q)U(w_h^{k_D(1)})]-G(e_1)\\ &\geqq\sum \pi_h(e_1)[qU(x_h^\varepsilon)+(1-q)U(w_h^{k_D(2)})]-G(e_1),\end{aligned}$$

where the second inequality follows from the fact that v^{1*} and $w^{k_D(1)}$ are most preferred schemes for a type e_1 agent in C^ε and D, respectively.

But we know that stochastic schemes are incentive-inefficient, and therefore must give less expected utility to the agent than the deterministic incentive scheme y^ε in Programme $P(\varepsilon)$. Intuitively, extraneous uncertainty has been created by the principal's randomization. Since the agent is risk averse, this can only diminish welfare. An analogous argument applies to equilibria in which the agent chooses actions randomly; when the principal randomizes over the choice of contracts, the agent's expected utility must be less than $\kappa(p,\varepsilon)$. To conclude, in subgame C^ε, the agent's equilibrium expected utility is at most $\gamma(\varepsilon)$.

There may be many equilibria in subgame C^ε. For example, when $\kappa(p,\varepsilon)\geqq\varepsilon$, in one equilibrium the agent chooses e_2 with probability $p<1$ and proposes $\{z^1,z^2\}$, which will be accepted by the principal. The agent's choice can be supported as an equilibrium move. If he proposes any other contract, the principal believes with probability one that action e_1 has been taken. Thus, the agent cannot gain by deviating. But beliefs in this equilibrium fail the refinement. Our main result is that in subgame C^ε in any equilibrium satisfying the refinement the agent always takes action e_2 and is compensated according to y^ε.

Proposition 3. *There exists a perfect-Bayesian equilibrium satisfying the refinement in subgame C^ε with the following allocation: the agent always take e_2 and is compensated according to y^ε; the principal's and the agent's equilibrium utilities are $\delta(\varepsilon)$ and $\gamma(\varepsilon)$, respectively. Moreover, any perfect-Bayesian equilibrium satisfying the refinement in subgame C^ε gives rise to this allocation.*

Proof. We begin by constructing such an equilibrium in subgame C^ε. Consider a contract $D=\{w^1,\ldots,w^K\}$; adopt the definition of $k_D(i)$ as in the proof of Proposition 1. The agent's equilibrium strategy in subgame C^ε is to take action e_2 and propose a contract D^ε with a single scheme y^ε: $D^\varepsilon=\{y^\varepsilon\}$. If the agent is to choose from contract C^ε, he picks v^{1*} [resp. x^ε] if he is type e_1 [resp. e_2]. If the agent is to choose from any other contract such as D, he picks $w^{k_D(i)}$ if he is type e_i. If the agent should take action e_1, he would propose $D=\{v^{1*}\}$.

The principal's equilibrium strategy is to accept the contract D^ε. On receiving a proposal D from the agent, the principal believes that action e_2 has been taken if it is a best action given D. Otherwise, she believes that action e_1 has been taken. Given the agent's strategy and the principal's beliefs about the agent's action, the principal accepts the agent's contract if her expected utility from any proposal is at least that from C^ε. Clearly, the equilibrium play results in an allocation described by the proposition.

It is easy to verify that these strategies constitute a perfect-Bayesian equilibrium in subgame C^ε. First, given the agent's strategy and the principal's belief, the principal's strategy is optimal. Constraint (12) and (y_h^ε) being a solution to Programme $P(\varepsilon)$ imply that given e_2, the principal is indifferent between D^ε and C^ε. Hence, accepting D^ε is optimal. Moreover, the principal's beliefs pass the refinement. Action e_2 is the unique best action given C^ε. If action e_2 is also the unique best action given a counteroffer D, according to the refinement, the principal believes that the action is e_2. This is what the equilibrium prescribes. Given the principal's strategy, the agent's strategy achieves his maximum equilibrium expected $\gamma(\varepsilon)$ in subgame C^ε, and therefore is a best response.

It remains to be proven that with the refinement imposed, any perfect-Bayesian equilibrium yields an allocation described in the proposition. We first demonstrate that the agent can get as close to $\gamma(\varepsilon)$ as he wishes. Let $\varepsilon'>0$ and consider scheme $(z^{\varepsilon' h})$ that solves the following programme: Choose a_h, $h=1,\ldots,n$, to

$$\text{maximize} \sum \pi_h(e_2)U(a_h)-G(e_2)$$

subject to

$$\sum \pi_h(e_2)U(a_h)-G(e_2)\geqq\sum \pi_h(e_1)U(a_h)-G(e_1)+\varepsilon', \tag{16}$$

$$\sum \pi_h(e_2)a_h\leqq\sum \pi_h(e_2)x_h^\varepsilon-\varepsilon'. \tag{17}$$

Suppose the agent offer $H^{\varepsilon'}=\{(z_h^{\varepsilon'})\}$ against C^ε. From (16), for any $\varepsilon'>0$, the agent's unique best action given $H^{\varepsilon'}$ is e_2. Given C^ε the agent's unique best action is also e_2. Thus the refinement says that on observing $H^{\varepsilon'}$ in subgame C^ε the principal must believe that the agent has chosen e_2. Constraint (17) guarantees that given e_2 the principal strictly prefers accepting $H^{\varepsilon'}$. Hence the principal must accept $H^{\varepsilon'}$. Since for $\varepsilon'=0$, (16) and (17) are identical to (11) and (12), respectively, the agent's expected utility from $z^{\varepsilon'}$ must tend to $\gamma(\varepsilon)$ as ε' tends to zero.

We now claim that in subgame C^ε in any equilibrium satisfying the refinement, the agent's equilibrium expected utility must be $\gamma(\varepsilon)$. To the contrary, suppose that there is an equilibrium in which the agent gets $\ell<\gamma(\varepsilon)$. For strictly positive and sufficiently small ε' the agent can guarantee a utility strictly bigger than ℓ by proposing $H^{\varepsilon'}$. This contradicts the assumption that the agent gets less than $\gamma(\varepsilon)$ in an equilibrium. Because y^ε is the unique solution to Programme $P(\varepsilon)$, the agent must take action e_2 and be rewarded according to y^ε. We conclude that in subgame C^ε the allocation in any equilibrium satisfying the refinement must be the one described in the proposition. ‖

In Proposition 3 we show how the principal can uniquely implement action e_2 at an arbitrarily small extra cost, provided perfect-Bayesian equilibria satisfy the refinement. As the paragraph preceding the proposition suggest, the refinement is necessary for the result.[13]

We can now prove that when perfect-Bayesian equilibria fulfill the refinement, equilibrium allocations must be second best. It is obvious that the principal's equilibrium payoff cannot be higher than β, her second-best expeted utility. Suppose in an equilibrium in the contract renegotiation game the principal gets $\eta < \beta$. Then there exists $\varepsilon > 0$ and sufficiently small such that in subgame C^ε her unique payoff in perfect-Bayesian equilibria satisfying the refinement is $\delta(\varepsilon) > \eta$. Moreover, the agent must accept the contract C^ε since in perfect-Bayesian equilibria of subgame C^ε, he obtains more than his reservation utility. This contradicts the assumption that η is an equilibrium payoff. Since $[e_2; v^{2*}]$ is the unique allocation that gives the principal an expected utility β, this must also be the allocation in every perfect-Bayesian equilibrium satisfying the refinement.

Theorem 1. *In the contract renegotiation game where the agent makes renegotiation offers, any perfect-Bayesian equilibrium satisfying the refinement leads to a second-best allocation, i.e., an equilibrium allocation in the standard principal-agent model with contract renegotiation disallowed.*

In this paper the analysis concentrates on a particular subgame C^0. We show that it has a second-best equilibrium. In order to prove the theorem, we then analyse equilibria when C^0 is slightly perturbed to C^ε. As we have discussed in the introduction, Matthews (1993) studies the same problem, but a contract in his model is simply a single sharing rule, not a menu. Thus subgames such as C^0 and C^ε in our model are excluded. Matthews (1993) proves a similar result: when selling the entire production process to the agent at a sufficiently high price is feasible, second-best allocations are the only equilibrium allocations, provided perfect-Bayesian equilibria pass a belief restriction (which is different from ours). Matthews's results imply that if a sales contract is feasible, there exist ex ante efficient equilibria in our game different from those we have analysed.[14]

We now discuss how our results can be generalized to the case of an arbitrary, finite number of actions. Suppose there are m actions: $e_1, \ldots, e_i, \ldots, e_m$, and each of these induces a distinct probability distribution $\Pi(e_i)$ on the revenue vector $(R_1, \ldots, R_n)$. The utility costs of these actions to the agent are given by $G(e_i)$. We adopt:

Assumption. Each action, e_i, $i = 1, \ldots, m$, is "strictly" implementable.[15]
That is, for each e_i, there is $\delta > 0$ such that for all $0 < \varepsilon < \delta$, there exists $(v_1^i, \ldots, v_h^i, \ldots, v_n^i)$ such that

$$\sum \pi_h(e_i)U(v_h^i) - G(e_i) \geqq \sum \pi_h(e_j)U(v_h^i) - G(e_j) + \varepsilon, \qquad j \neq i. \tag{18}$$

An incentive-efficient for action e_i is a solution to the following programme: Choose v_h, $h = 1, \ldots, n$, to

$$\text{minimize} \sum \pi_h(e_i)v_h$$

13. Proposition 3 contrasts with a result in the informed-principal model. Proposition 11 in Maskin and Tirole (1992) demonstrates that refinements such as Cho–Kreps, and Farrell–Grossman–Perry may rule out interim inefficient equilibria in that model. Our refinement is weaker since it only uses a dominance argument, and it rules out *ex ante* inefficient equilibria.

14. Since a contract may consist of a single scheme, equilibria in Matthews's model are equilibria in ours. A sales contract may not be feasible if the domain of the agent's monetary utility function U is bounded below.

15. Proposition 2 and Corollary 2 in Hermalin and Katz (1991) characterize conditions on $\Pi(e_i)$ and $G(e_i)$ for which the set of inequalities in (18) admits a solution when ε is set at zero.

subject to

$$\sum \pi_h(e_i)U(v_h) - G(e_i) \geqq \sum \pi_h(e_j)U(v_h) - G(e_j), \qquad j=1,\ldots,m, \tag{19}$$

$$\sum \pi_h(e_i)U(v_h) - G(e_i) \geqq 0. \tag{20}$$

Let $v^{i*} = (v_1^{i*}, \ldots, v_n^{i*})$ be an incentive-efficient scheme for action e_i. It is straightforward to verify that inequality (20) must bind. Consider the set of incentive-efficient schemes, $V^* = \{v^{1*}, \ldots, v^{i*}, \ldots, v^{m*}\}$. Then v^{i*} is a most preferred scheme in for a type e_i agent in V^*. If he has taken action e_i, we know from (20) that if he picks v^{i*}, his expected utility is zero. But none of the other schemes offer a higher expected utility. Consider scheme v^{j*}, $j \neq i$. Given v^{j*}, e_j is a best action, and from constraints (19), and constraint (20) being binding, we see that

$$\sum \pi_h(e_i)U(v_h^{j*}) - G(e_i) \leqq 0.$$

A second-best action is given by e_s where

$$\sum \pi_h(e_s)(R_h - v_h^{s*}) \geqq \sum \pi_h(e_i)(R_h - v_h^{i*}), \qquad i=1,\ldots,m.$$

A second-best allocation is defined by $[e_s; v^{s*}]$. For any given set of parameters, there may be many second-best allocations. Nevertheless, our results apply to every second-best allocation. Proposition 1 generalizes easily. According to Proposition 0, $\{v^{1*}, \ldots, v^{i*}, \ldots, v^{m*}\}$ is incentive-efficient for any random action. Consider the subgame defined by the contract $\{v^{1*}, \ldots, v^{i*}, \ldots, v^{m*}\} \equiv C^0$. Any distribution on the agent's action can be an equilibrium action in subgame C^0.

The construction of these equilibria closely resembles those in Proposition 1. For any counter-proposal from the agent, $D = \{w^1, \ldots, w^K\}$, let $k_D(i)$ be the smallest index of a type e_i agent's most preferred schemes in D. For any D, a type e_i agent picks scheme $w^{k_D(i)}$. Let $\Psi(D)$ be the set of best actions with respect to D:

$$\Psi(D) = \{e_i : i \in \operatorname{argmax}_j \sum \pi_h(e_j)U(w_h^{k_D(j)}) - G(e_j), j=1,\ldots,m.\}.$$

The principal believes that the agent chooses his actions according an arbitrary distribution on $\Psi(D)$. She rejects D if and only if D gives her less expected utility than C^0 according to her beliefs. Using the argument in Proposition 1, one easily shows that any random action is an equilibrium action in C^0. Therefore, the seond-best allocation $[e_s; v^{s*}]$ is an equilibrium allocation.

To prove that every equilibrium allocation is second-best when equilibria satisfy the refinement, first consider a scheme $v^s(\varepsilon)$ that solves the following programme: for $\varepsilon > 0$, choose v_h, $h=1,\ldots,n$, to

$$\text{minimize} \sum \pi_h(e_s)v_h$$

subject to

$$\sum \pi_h(e_s)U(v_h) - G(e_s) \geqq \sum \pi_h(e_i)U(v_h) - G(e_i) + 2\varepsilon, \qquad i \neq s,$$

$$\sum \pi_h(e_s)U(v_h) - G(e_s) \geqq \varepsilon.$$

In this programme, the agent is given a strict incentive to perform e_s; what is more, a rent $\varepsilon > 0$ is guaranteed with this action.

Next, consider $i \neq s$, and for $\varepsilon_i > 0$, let $v^i(\varepsilon_i)$ be the solution to the following programme: choose v_h, $h=1,\ldots,n$, to

$$\text{minimize} \sum \pi_h(e_i)v_h$$

subject to

$$\sum \pi_h(e_i)U(v_h) - G(e_i) \geqq \sum \pi_h(e_j)U(v_h) - G(e_j) + \varepsilon_i, \qquad i \neq j,$$
$$\sum \pi_h(e_i)U(v_h) - G(e_i) \geqq 0.$$

Here, the agent strictly prefers to pick action e_i given the scheme, but no rent is available through $v^i(\varepsilon_i)$.

Furthermore, pick $\varepsilon_i > 0$ sufficiently small such that

$$\sum \pi_h(e_i)U(w_h^i) - G(e_i) < \varepsilon, \tag{21}$$

where w^i solves the following programme: choose v_h, $h = 1, \ldots, n$, to

$$\text{maximize} \sum \pi_h(e_i)U(v_h) - G(e_i)$$

subject to

$$\sum \pi_h(e_i)U(v_h) - G(e_i) \geqq \sum \pi_h(e_j)U(v_h) - G(e_j), \qquad i \neq j,$$
$$\sum \pi_h(e_i)(R_h - v_h^i(\varepsilon_i)) \geqq \sum \pi_h(e_i)(R_h - v_h).$$

That is, w^i represents the incentive-efficient scheme given e_i and status quo scheme $v^i(\varepsilon_i)$ for the principal. Although w^i may relax all the slacks in the incentive constraints, by inequality (21) the original slack ε_i is so small that the agent's gain in expected utility must be less than ε. For any $\varepsilon > 0$, and for each $i \neq s$, an ε_i that satisfies the above conditions must exist. To see this, suppose $\varepsilon_i = 0$, then $v^i(\varepsilon_i)$ is simply an incentive-efficient contract for action e_i and $v^i(\varepsilon_i) = w^i$, hence (21) is satisfied automatically. By continuity, for ε_i sufficiently close to 0, inequality (21) will still be true.

Now consider a perturbation C^0: $C^\varepsilon = \{v^1(\varepsilon_1), \ldots, v^i(\varepsilon_i), \ldots, v^s(\varepsilon), \ldots, v^m(\varepsilon_m)\}$. Clearly, each ε_i, $i \neq s$, must go to zero as ε goes to zero, and therefore C^ε tends to C^0 as well. Thanks to $\varepsilon_i > 0$ in the relaxation of the incentive constraints in the programme for $v^i(\varepsilon_i)$, we can easily verify that a type e_i agent will strictly prefer to pick scheme $v^i(\varepsilon_i)$ from C^ε. Similarly, a type e_s agent will strictly prefer to pick scheme $v^s(\varepsilon)$. Note that for the case of only two actions in the earlier analysis, ε_1 can simply be taken as zero, since the incentive constraint in Programme $P(e_1)$ does not bind.

For $i \neq s$, although each of the incentive constraints for the programme that determines $v^i(\varepsilon_i)$ has been slightly relaxed, the amount of slack ε_i has been chosen so small that even if the agents is allowed to propose an alternative incentive-efficient scheme that completely relaxes this slack (such as w^i), he will gain by strictly less than ε. Put differently, although the scheme $v^i(\varepsilon_i)$ allows a slack to ensure that type e_i must choose $v^i(\varepsilon_i)$ in contract C^ε, this slack is so small relative to ε in the programme for $v^s(\varepsilon)$ that $v^i(\varepsilon_i)$ is almost incentive-efficient for action e_i.

Finally, for the programme that determines $v^s(\varepsilon)$, *both* the incentive and reservation utility constraints have been relaxed. Given $v^s(\varepsilon)$, the agent obtains ε by performing e_s. The important point, however, is that, by the construction of $v^i(\varepsilon_i)$'s, none of the other actions can allow the agent to achieve ε.

Observe that the agent cannot gain by randomizing between actions; if he did randomize over actions, then the principal's status quo utility at renegotiation would be determined by schemes with different amounts of slacks, and therefore by schemes with smaller amounts of slacks than that in $v^s(\varepsilon)$. This implies that the agent's equilibrium expected would not be the maximum of all possible equilibria. Also, the agent cannot gain if the principal randomizes between the contract C^ε and his proposal because extraneous uncertainty is introduced.

In an equilibrium in subgame C^{ε}, the agent must choose e_s with strictly positive probability, an extension of Proposition 2. Although contract C^{ε} is incentive-inefficient generally and thus may be renegotiated to an incentive-efficient contract even if the agent does not take action e_s, the agent cannot obtain ε if he never takes action e_s. The maximum slack in C^{ε} resides in $v^s(\varepsilon)$. Thus, the maximum equilibrium expected utility the agent can obtain through renegotiation is by taking action e_s and proposing an incentive-efficient contract that relaxes the maximum slack in C^{ε}. Furthermore, the refinement guarantees that if the agent offers a contract with a single scheme which solves a programme similar to $P(\varepsilon)$, the principal must accept it, allowing the agent to achieve his maximum expected utility. By perturbing the contract C^0 slightly, the principal achieves the second best approximately. It follows that in an equilibrium satisfying the refinement, the equilibrium allocation must be second best.

5. CONCLUSIONS

In this paper we have analysed contract renegotiation in a principal-agent model. Contract renegotiation offers are made by the agent. A refinement on the principal's beliefs is imposed: if exactly one action is optimal with respect to both the principal's and the agent's contract, she believes that the agent has chosen it. With this refinement, perfect-Bayesian equilibria give rise to allocations identical to the second best in the standard principal-agent model without renegotiation. When renegotiation is led by the agent and when equilibria satisfy the refinement, equilibrium allocations are ex ante efficient.

In the paper the principal is assumed risk neutral. This assumption is unimportant. The theorem is true when the principal is risk averse. None of our results depends on the particular insurance characteristic of incentive schemes. If the principal is risk averse, we can simply replace the principal's objective function in each optimization programme in this paper by her utility function.

The results in Fudenberg and Tirole (1990) and here suggest that equilibrium allocations depend critically on the exact renegotiation mechanism. Future research to address robustness issues seems necessary. It would be interesting to study game forms that permits renegotiation offers to be made by both the principal and the agent. As an example, the agent and the principal can be chosen to be the proposer with probabilities θ and $1-\theta$, respectively. Then the second best is not an equilibrium allocation for θ strictly less than one. This is because the principal must offer a full insurance scheme at renegotiation if she believes that a costly action has been taken with probability one. Because of this, the agent will be rewarded according to a full insurance scheme with positive probability. So the equilibrium allocation cannot be second best. However, we speculate that as θ approaches one, there exist approximately second-best equilibria. The formal characterization of equilibria and whether equilibrium refinements can rule out inefficient ones await research.

The renegotiation models in Fudenberg and Tirole (1990) and in this paper use a static production technology: revenues are realized only once. In dynamic models, the agent's action may induce revenues over many periods. Then there are many possible dates for renegotiation. If renegotiation can only take place before any revenue is realized and is led by the agent, results in this paper should apply. Analysis of the equilibria of dynamic models when agent-led renegotiation occurs after some revenues have realized requires further research.

APPENDIX

Proof of Proposition 0. We begin by rewriting the programme for the optimal contract for action e_2 with probability p: Choose v_h^i, $h=1,\ldots,n$, $i=1,2$, to

$$\text{minimize} \sum (1-p)\ \pi_h(e_1)v_h^1 + \sum p\pi_h(e_2)v_h^2$$

subject to

$$\sum \pi_h(e_1)U(v_h^1) - G(e_1) \geqq \sum \pi_h(e_2)U(v_h^2) - G(e_2), \tag{22}$$

$$\sum \pi_h(e_2)U(v_h^2) - G(e_2) \geqq \sum \pi_h(e_1)U(v_h^1) - G(e_1), \tag{23}$$

$$\sum \pi_h(e_1)U(v_h^1) - G(e_1) \geqq 0, \tag{24}$$

$$\sum \pi_h(e_2)U(v_h^2) - G(e_2) \geqq 0, \tag{25}$$

$$\sum \pi_h(e_1)U(v_h^1) \geqq \sum \pi_h(e_1)U(v_h^2), \tag{26}$$

$$\sum \pi_h(e_2)U(v_h^2) \geqq \sum \pi_h(e_2)U(v_h^1). \tag{27}$$

Consider a relaxed programme in which (22), (23), (26) and (27) are replaced by

$$\sum \pi_h(e_1)U(v_h^1) - G(e_1) \geqq \sum \pi_h(e_2)U(v_h^1) - G(e_2), \tag{28}$$

$$\sum \pi_h(e_2)U(v_h^2) - G(e_2) \geqq \sum \pi_h(e_1)U(v_h^2) - G(e_1). \tag{29}$$

Clearly, (28) is implied by (22) and (27), and (29) by (23) and (26). In the relaxed programme, the constraints are (24), (25), (28) and (29). Notice that (28) and (24) [resp. (29) and (25)] are the incentive and reservation utility constraints in the programme for the incentive-efficient scheme for action e_1 [resp. e_2]. Moreover the objective fuction of the relaxed programme is a weighted average of the objective functions of the programmes for the incentive-efficient contracts for actions e_1 and e_2. Therefore x^{1*} and x^{2*} solve the relaxed programme if and only if they are the incentive-efficient schemes for e_1 and e_2, namely v^{1*} and v^{2*}.

Finally, we show that the solution to the relaxed programme satisfies the constraints of the original programme. At a solution to the relaxed programme (24) and (25) must hold as equalities. Therefore (22) and (23) are true. Constraints (24) and (25) holding as equalities and (29) together imply that (26) holds. A similar argument establishes that (27) is also satisfied. ||

Acknowledgements. Part of this research was carried out during a visit at the Department of Economics, University of Bonn; their hospitality is acknowledged. My thanks to Douglas Gale, Jacob Glazer, Michael Riordan, Robert Rosenthal, two referees, and especially Steven Matthews for their helpful advice. Seminar participants at Boston, Brown and Northwestern Universities, Stanford Institute of Economic Theory, and the 1991 North American Summer Meeting of the Econometric Society at the University of Pennsylvania gave me useful comments.

REFERENCES

ARNOTT, R. AND STIGLITZ, J. E. (1988), "Randomization with Asymmetric Information", *Rand Journal of Economics*, **19**, 344–362.

FUDENBERG, D. AND TIROLE, J. (1988), "Moral Hazard and Renegotiation in Agency Contracts" (Working Paper No. 494, MIT).

FUDENBERG, D. AND TIROLE, J. (1990), "Moral Hazard and Renegotiation in Agency Contracts", *Econometrica*, **58**, 1279–1319.

FUDENBERG, D AND TIROLE, J. (1991) *Game Theory* (Cambridge, Mass.: MIT Press).

GROSSMAN, S. AND HART, O. D. (1983), "An Analysis of the Principal-Agent Problem", *Econometrica*, **51**, 7–45.

HERMALIN, B. E. AND KATZ, M. L. (1991), "Moral Hazard and Verifiability: The Effects of Renegotiation in Agency", *Econometrica*, **59**, 1735–1754.

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

HOLMSTROM, B. AND MYERSON, R. B. (1983), "Efficient and Durable Decision Rules with Incomplete Information", *Econometrica*, **51**, 1799–1819.

KOHLBERG, E. AND MERTENS, J.-F. (1986), "On the Strategic Stability of Equilibria", *Econometrica*, **54**, 1003–1037.

MA, C.-T. A. (1991), "Adverse Selection in Dynamic Moral Hazard", *Quarterly Journal of Economics*, **106**, 255–275.

MASKIN, E. AND TIROLE, J. (1992), "The Principal-Agent Relationship with an Informed Principal, II: Common Values", *Econometrica*, **60**, 1–42.

MATTHEWS, S. A. (1993), "Renegotiation of Sales Contracts under Moral Hazard" (Discussion Paper No. 1051, Center for Mathematical Studies in Economics and Management Science, Northwestern University).