Moral Hazard and Renegotiation in Agency Contracts

Drew Fudenberg, Jean Tirole

Econometrica, Volume 58, Issue 6 (Nov., 1990), 1279-1319.

Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199011%2958%3A6%3C1279%3AMHARIA%3E2.0.CO%3B2-E

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Econometrica is published by The Econometric Society. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/econosoc.html.

Econometrica
©1990 The Econometric Society

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2002 JSTOR

http://www.jstor.org/
Tue Feb 5 11:51:07 2002

Econometrica, Vol. 58, No. 6 (November, 1990), 1279–1319

MORAL HAZARD AND RENEGOTIATION IN AGENCY CONTRACTS

BY DREW FUDENBERG AND JEAN TIROLE[1]

Previous analyses of principal-agent problems with moral hazard assume the parties can commit to a contract that will not be renegotiated. We allow the contract to be renegotiated after the agent's choice of action and before the observation of the action's consequences. Here, the principal cannot induce the agent to take a high level of effort with probability one, since the principal would renegotiate the contract to shield the agent from risk, and in equilibrium the agent randomizes over effort levels. The optimal contract gives the agent a menu of compensation schemes: safe ones intended for low-effort workers, and risky ones for those whose effort is high. The optimal contract may give the agent a positive rent, in contrast to the case without renegotiation.

1. INTRODUCTION

CONSIDER THE STANDARD MORAL HAZARD MODEL of designing a contract between a risk-averse agent and a risk-neutral principal when the agent's action is subject to moral hazard. Previous analyses (e.g. Holmström (1979), Shavell (1979)) assume that the parties can commit themselves to a contract that will not be renegotiated. While such commitment is likely to be credible in some situations, in others it may not be, especially if there are long lags between the agent's choice of action and the time when all of the (stochastic) consequences of that action will have been revealed. In this case the parties may be able to renegotiate in the "interim" phase between the agent's action and the observation of its consequences. This paper studies the implications of the renegotiation for the structure of optimal contracts.[2]

Fudenberg, Holmström, and Milgrom (1990) show that the possibility of renegotiation has no impact if the parties know each other's preferences over contracts at every potential recontracting date and the agent has unrestricted access to a perfect capital market. However, the former condition is unlikely to be satisfied if the agent's actions correspond to investment decisions with long-term consequences. Imagine for example that the action the agent takes today will influence the likely course of events for the next five years. The standard informativeness argument (Shavell (1979) and Holmström (1979)) suggests that the agent's compensation should then depend on the outcomes to be observed five years hence. However, since only the agent knows which action

[1] Research support from National Science Foundation Grants SES 8808204 and SES 8520837 and the Taussig Visiting Professorship at Harvard is gratefully acknowledged. We would like to thank Bob Gibbons, Paul Healy, Oliver Hart, Jean-Jacques Laffont, Mike Whinston, a co-editor, and a referee for helpful comments, and Dimitrios Vayanos for able research assistance.

[2] Recent studies of renegotiation in a Nash implementation context include Aghion-Dewatripont-Rey (1989), Green-Laffont (1987), Hart-Moore (1988), Maskin-Moore (1987), and, in an adverse-selection context, Dewatripont (1989), Hart-Tirole (1988), and Laffont-Tirole (1988).

was chosen, throughout the intervening years the agent has private information about the probability distribution of future outcomes. If the parties have the opportunity to renegotiate the contract during this intervening period, they might well choose to do so.

We analyze the renegotiation problem in the following simple model. First, the parties meet and sign an original or *ex-ante* contract c_1 which specifies the agent's monetary compensation as a function of the realized outcome. Then the agent chooses an effort level e. This effort generates a probability distribution $p(e)$ over outcomes; the principal will observe the realized outcome but not the agent's choice of e. After the effort is chosen, but before the outcome is realized, the parties have the opportunity to renegotiate, replacing c_1 by another contract c_2. Here we must specify the way in which renegotiation proceeds. We examine the case in which the principal is able to implement the optimal mechanism at the renegotiation stage, which will typically involve the principal offering the agent a menu of contracts, one for each level of effort that the agent may have chosen. This specification is comparatively simple because the principal has no private information. The results of Maskin-Tirole (1988) on the informed principal problem show that the same conclusions are obtained if the agent is the one who proposes contracts during the renegotiation and one requires the contract to be "strongly renegotiation-proof."[3]

It is easy to see that the equilibrium of the renegotiation model generally differs from that when the parties can commit not to renegotiate. The full-commitment contract will generally expose the agent to some risk in order to induce him to choose the desired level of effort e^*, but once the effort has been made, it would be more efficient to provide the agent with complete insurance. If the principal were certain that the agent chose e^*, he would then offer the agent a sure payment which yields the same expected utility on the assumption that e^* was chosen. Foreseeing that his eventual payment would be independent of the outcome, the agent would then prefer to choose the lowest feasible level of effort $\underline{e}$ (unless $e^* = \underline{e}$). This suggests that, if renegotiation is possible, the equilibrium will be in mixed strategies: In order to make credible a contract that induces the agent to choose some $\hat{e} > \underline{e}$ by promising a higher payment following good outcomes, the principal must also induce the agent to choose lower effort levels with sufficiently high probability that the contract would not be renegotiated.

The idea here can be grasped from Stiglitz's (1977) result that in a model of insurance under adverse selection, the optimal contract for a monopolist won't offer much insurance to the "good" type if the probability of the "bad" type is sufficiently high. The analogy with Stiglitz's model is that our moral-hazard problem becomes an adverse-selection problem at the renegotiation stage, where the agent's "type" corresponds to his level of effort. A high-effort type faces a higher probability of a good outcome than a low-effort type does. The

[3] A contract is strongly renegotiation-proof if it is not renegotiated in any equilibrium of the renegotiation stage.

principal's objective at the renegotiation stage is to find the profit-maximizing insurance policies or compensation schemes to offer an agent who has private information about his risk level. As in Stiglitz, the fact that the different types of agent have different preferences over insurance policies means that the principal can gain by offering a menu of compensation schemes. The difference with the Stiglitz model is that the risks he considers are exogenous, while in our case the risks the agent faces at the renegotiation stage stem from the endogenously determined original contract.

From previous work on renegotiation, we expect that any outcome the principal can obtain by some choice of contract can be attained by a contract that is renegotiation-proof, i.e., will not be renegotiated at the interim stage. We verify that this is so at the beginning of Section 2, and proceed to characterize the set of renegotiation-proof contracts.

Section 3 supposes that there are two possible outcomes, and two possible efforts for the agent. We first derive necessary conditions for a contract to be renegotiation-proof, and then show that these conditions are sufficient if the probability of high effort is sufficiently small.

Section 3A then examines the principal's optimal choice of *ex-ante* contract. We show that for a range of utility functions for the agent, including exponential and logarithmic forms, any distribution that can be implemented by a renegotiation-proof contract can be implemented with zero *ex-ante* rent for the agent. It is easy to show that this implies that the cost-minimizing renegotiation-proof contract for a given distribution is the same as the cost-minimizing contract under commitment, if this distribution can be implemented at all. Thus, the force of the renegotiation-proof constraint is not to change the way that given distributions are implemented, but only to change which distributions are feasible.

Interestingly, though, there are some utility functions for which the set of renegotiation-proof distributions grows as the agent's *ex-ante* rent increases. In this case the principal may prefer to increase the agent's compensation in order to relax the renegotiation-proofness constraint, so that the optimal contract may differ from that under commitment not only in the choice of distribution but also in the way that distribution is implemented.

Our analysis of renegotiation-proof contracts assumes that when the agent is indifferent between several actions he is willing to randomize between them in the way that the principal most prefers. Section 3B shows that this assumption is not necessary: There are initial contracts the principal can offer such that the unique continuation equilibrium (which involves renegotiation) gives the principal the maximal payoff from a renegotiation-proof contract. Section 4 develops the link with the commitment model.

Section 5A extends the model to a continuum of efforts. The optimal renegotiation-proof contract induces a continuous distribution of effort levels between the lowest possible level and one that exceeds the optimal commitment effort. This distribution is given by a generalized hazard rate condition that reflects the tradeoff at the interim stage between the value of increased

insurance for an agent who has exerted a given level of effort e and the cost, imposed by interim incentive compatibility, of increasing the interim rent for lower levels of effort. Appendix 2 extends the results to a continuum of outcomes as well.

Section 5B shows that the optimal renegotiation-proof contract, in which the agent chooses from a menu of incentive schemes, can alternatively be implemented by offering a single *ex-ante* contract which is later renegotiated towards more insurance. This single incentive scheme is the incentive scheme in the renegotiation-proof menu that corresponds to the highest equilibrium effort.

As discussed in the conclusion (Section 6), our theory may shed some light on why compensation of managers and contractors is frequently insensitive to the information obtained after the relationship is terminated, and why executives have considerable discretion to adjust the riskiness of their compensation during their tenure within the firm.

2. RENEGOTIATION WITH TWO EFFORT LEVELS AND TWO OUTCOMES

2A. *The Model*

We begin by analyzing a simple model in which the agent has only two levels of effort. Section 5 considers the case of a continuum of effort levels, and finds that the main results of this section extend in a natural way. We assume that the agent's utility function for income w and effort e is additively separable, $V(w,e) = U(w) - D(e)$, and we normalize the agent's utility of not working for the principal to zero. We assume that $U' > 0$ and $U'' < 0$ (i.e. the agent likes income and is risk averse) and let $\Phi(U)$ be the inverse function corresponding to U.

We assume that the domain of U is $(-\infty, +\infty)$, that $\lim_{w \to -\infty} U(w) = -\infty$, and that $\lim_{w \to +\infty} U(w) = +\infty$. (Except for Section 3B, our results extend to the case where utility is bounded below, which allows limited liability. What matters for our proofs is that wages can be set arbitrarily and that utilities are unbounded above.) There are two possible outcomes/incomes for the principal, g and b; $g > b$, the probability of outcome g when the agent chooses effort e is denoted $p(e)$. In this section we further assume that there are only two levels of effort, $\underline{e}$ and $\bar{e}$, with $D(\underline{e}) < D(\bar{e})$ and $p(\underline{e}) < p(\bar{e})$. The principal is assumed to be risk-neutral; her objective is to maximize the difference between her expected revenue $I(e) = p(e)g + (1 - p(e))b$ and the expected wage bill $E(w|e)$. The principal has the option of not employing the agent; we normalize this shut-down income to be zero.

A *feasible contract* c is a pair of compensation schemes $\{c(\underline{e}), c(\bar{e})\}$ from which the agent chooses before the outcome is realized. A *feasible compensation scheme* $c(e)$ specifies two utility levels $\{U_g(e), U_b(e)\}$, meaning that the principal will pay the agent $w_g(e) = \Phi(U_g(e))$ if outcome g occurs, and $w_b(e) = \Phi(U_b(e))$ following outcome b. Let C denote the set of feasible contracts.

The *extensive form* of the game is as follows. The principal offers an "initial contract" $c_1 \in C$. The agent accepts or rejects the contract. In case of rejection the agent does not work for the principal and both get their reservation utility,

which is normalized to zero. If the agent accepts, he chooses a probability distribution $x = x(c_1)$ over the effort levels $\{\underline{e}, \bar{e}\}$, with $x = \text{Prob}(e = \bar{e})$. The principal does not observe x or the realized effort level e.

After the choice of effort is realized and before the outcome is observed, the principal offers the agent a new contract $c_2 \in C$. This is the "interim stage." If c_2 is rejected, c_1 is the "final contract" and determines the agent's compensation. If c_2 is accepted, it becomes the final contract. Then the agent chooses an element $c(e)$ from the relevant menu of compensation schemes. (This is equivalent to making an announcement $\tilde{e}$ of his effort level.) Last, the outcome is realized and the wage paid to the agent.

We will discuss the choice of the extensive form in the conclusion. For the moment, let us note that C is the relevant contract space. On the one hand, the principal must be allowed to offer contracts where $c(\underline{e}) \neq c(\bar{e})$, because he does not know the agent's choice of effort, and so he may want to screen the agent's interim preferences through a menu of compensation schemes. On the other hand, there is no loss of generality in restricting the contract space to be C. The revelation principle implies that, at the interim stage, the principal can implement any allocation obtained through a complex contract (a general game form where the wage depends on a sequence of announcements by both parties) by using a direct revelation mechanism, in which the agent announces his past level of effort (type). It also implies that we can require that the agent report his (already chosen) effort level truthfully at the interim stage.

Section 3A derives an upper bound on the principal's payoff in all Nash equilibria. Section 3B shows that there exists a Nash equilibrium that is perfect Bayesian and yields this upper bound. (We will use the weakest version of perfect Bayesian equilibrium (PBE), which requires that the strategies be sequentially rational given the players' beliefs, and that the beliefs be obtained wherever possible from observed actions and equilibrium strategies using Bayes rule. Sequential equilibrium (Kreps-Wilson (1982)) cannot be applied to this model due to the continuum of wage offers, but in a discrete version of the game, the two would be equivalent.)

2B. *Renegotiation-proof Contracts*

A contract is renegotiation-proof if the principal will not choose to alter it at the renegotiation stage, given his beliefs about the distribution x of the agent's effort. Proposition 2.1 below shows that there is no loss of generality in restricting attention to such contracts. For this reason, it is interesting to characterize the renegotiation-proof contracts in detail.

DEFINITION 2.1: A contract $c = \{U_g(e), U_b(e)\}_{e=\underline{e},\bar{e}}$ is a *renegotiation-proof contract for distribution* x if it minimizes the expected compensation cost given distribution x:

$$M(\tilde{c}, x) = x\left[p(\bar{e})\Phi\left(\tilde{U}_g(\bar{e})\right) + (1 - p(\bar{e}))\Phi\left(\tilde{U}_b(\bar{e})\right)\right] + (1 - x)\left[p(\underline{e})\Phi\left(\tilde{U}_g(\underline{e})\right) + (1 - p(\underline{e}))\Phi\left(\tilde{U}_b(\underline{e})\right)\right]$$

among the feasible contracts $\tilde{c} = \{\tilde{U}_g(e), \tilde{U}_b(e)\}_{e=\underline{e},\bar{e}}$ that satisfy for all e:

(i) IIC(e) (Interim Incentive Compatibility for type e):

$$p(e)\tilde{U}_g(e) + (1-p(e))\tilde{U}_b(e) \geqslant p(e)\tilde{U}_g(\tilde{e}) + (1-p(e))\tilde{U}_b(\tilde{e}) \quad \forall \tilde{e}.$$

(ii) IIR(e) (Interim Individual Rationality for type e):

$$p(e)\tilde{U}_g(e) + (1-p(e))\tilde{U}_b(e) \geqslant p(e)U_g(e) + (1-p(e))U_b(e).$$

Note that at the interim stage the agent's effort has already been chosen. This is why neither the principal's expected income nor the agent's disutility of effort enter Definition 2.1. Note also that there are two individual rationality constraints, one for each type.

Finally, note that we consider the renegotiation-proofness of arbitrary pairs (c, x). We do not require that distribution x be consistent with the agent's incentives to choose effort at the *ex-ante* stage, given he expects the contract c to remain in force. This is because we find it simpler to work backward from the end of the game tree, first defining renegotiation-proofness and then determining the agent's choice of actions.

PROPOSITION 2.1: *If there is a Nash equilibrium where the distribution over effort levels is* $(x^*, 1-x^*)$, *the initial contract is* $\hat{c}$ *and the final contract is* c^*, *there is an equilibrium with the same distribution over efforts where* c^* *is the initial contract as well as the final one.*

PROOF: Imagine that the principal offers $c_1 = c^*$ at the initial stage, and that the agent accepts, and chooses distribution $(x^*, 1-x^*)$. If there were a contract, $\tilde{c}$, that upset c^*, it would have to offer at least as much expected utility to every type of the agent, and a strictly higher payoff to the principal. But then the principal would have offered $\tilde{c}$ instead of c^* when renegotiating the contract $\hat{c}$. Thus c^* is renegotiation-proof for distribution x^*.

Moreover, since the agent's utility depends only on the final contract and not on the initial one, if it is an equilibrium to choose distribution x^* when $\hat{c}$ is the initial contract and c^* is the final one, it is also an equilibrium to choose distribution x^* when $c_1 = c_2 = c^*$. *Q.E.D.*

REMARK: While Proposition 2.1 is stated in the notation for the two-outcome, two-effort case, it clearly extends to arbitrary outcome and effort spaces.

LEMMA 2.1: *If* $c = \{U_g(e), U_b(e)\}_{e=\underline{e},\bar{e}}$ *is a renegotiation-proof contract for distribution* x, *then either*

$$p(\underline{e})U_g(\underline{e}) + (1-p(\underline{e}))U_b(\underline{e}) > p(\bar{e})U_g(\bar{e}) + (1-p(\bar{e}))U_b(\bar{e}),$$

so that type $\underline{e}$ has strictly higher interim utility then $\bar{e}$, or the following conditions all hold:

(a) $U_g(\underline{e}) = U_b(\underline{e}) = \underline{U}$,

(b) $U_g(\bar{e}) \geqslant U_b(\bar{e})$,

(c) $p(\underline{e})U_g(\underline{e}) + (1-p(\underline{e}))U_b(\underline{e}) = p(\underline{e})U_g(\bar{e}) + (1-p(\underline{e}))U_b(\bar{e})$,

so the interim incentive compatibility constraint is binding for type $\underline{e}$.

PROOF: Adding IIC($\underline{e}$) and IIC($\bar{e}$) yields "monotonicity":

$$\left(p(\bar{e}) - p(\underline{e})\right)\left(\left(\tilde{U}_g(\bar{e}) - \tilde{U}_b(\bar{e})\right) - \left(\tilde{U}_g(\underline{e}) - \tilde{U}_b(\underline{e})\right)\right) \geqslant 0.$$

The minimization of $M(\tilde{c}, x)$ in Definition 2.1 leads us to consider four cases depending on which interim incentive compatibility constraints are binding. First, if IIC($\underline{e}$) and IIC($\bar{e}$) are satisfied with equality, $\tilde{U}_g(\bar{e}) - \tilde{U}_b(\bar{e}) = \tilde{U}_g(\underline{e}) - \tilde{U}_b(\underline{e})$, and the two interim incentive constraints yield $\tilde{U}_g(\bar{e}) = \tilde{U}_b(\bar{e}) = \tilde{U}_g(\underline{e}) = \tilde{U}_b(\underline{e})$; that is, the contract is a single, full insurance compensation scheme. Second, if both IIC($\bar{e}$) and IIC($\underline{e}$) hold with strict inequality, then minimization of $M(\tilde{c}, x)$ implies that $\tilde{U}_g(\bar{e}) = \tilde{U}_b(\bar{e}) = \overline{U}$ and $\tilde{U}_g(\underline{e}) = \tilde{U}_b(\underline{e}) = \underline{U}$. This is only consistent with incentive compatibility if $\overline{U} = \underline{U}$, in which case the IIC constraints are only weakly satisfied, a contradiction. Third, assume that IIC($\bar{e}$) is satisfied with strict inequality and IIC($\underline{e}$) is satisfied with equality. The minimization of $M(\tilde{c}, x)$ subject to all constraints except IIC($\bar{e}$) yields $\tilde{U}_g(\underline{e}) = \tilde{U}_b(\underline{e}) = \underline{U}$; monotonicity and IIC($\bar{e}$) being satisfied with strict inequality then imply that $\tilde{U}_g(\bar{e}) > \tilde{U}_b(\bar{e})$. Fourth, suppose that IIC($\underline{e}$) is satisfied with strict inequality and IIC($\bar{e}$) is satisfied with equality. The minimization of $M(\tilde{c}, x)$ subject to all constraints except IIC($\underline{e}$) yields $\tilde{U}_g(\bar{e}) = \tilde{U}_b(\bar{e}) = \overline{U}$. Monotonicity and IIC($\underline{e}$) being satisfied with strict inequality then imply that $\tilde{U}_g(\underline{e}) < \tilde{U}_b(\underline{e})$ (that is, the bad type is rewarded more in the bad outcome). In this fourth case,

$$p(\underline{e})\tilde{U}_g(\underline{e}) + (1-p(\underline{e}))\tilde{U}_b(\underline{e}) > p(\underline{e})\tilde{U}_g(\bar{e}) + (1-p(\bar{e}))\tilde{U}_b(\bar{e}) = \overline{U},$$

so that the bad type's interim utility exceeds the good type's. *Q.E.D.*

Lemma 2.1 yields two potential classes of renegotiation-proof contracts, one in which the bad type's interim utility exceeds the good type's, and one satisfying conditions (a), (b), and (c) above.

In the first class of contracts, the roles of the good and bad types and the good and bad outcomes are reversed, as the "bad type" receives a higher wage when the bad outcome occurs. Clearly under such a contract the agent would never choose to work at the *ex-ante* stage. For this reason, we will focus on the second class of contracts from now on, with $U_g > U_b$ and so $w_g > w_b$.

The lemma shows that any renegotiation-proof contract where type $\bar{e}$'s interim utility is at least as high as type $\underline{e}$'s must satisfy conditions (a), (b), and (c). The intuition for this conclusion can be grasped from the familiar Rothschild-Stiglitz (1976) and Wilson (1977) depiction of the insurance problem in the wage rather than utility space. Figure 1 identifies compensation schemes

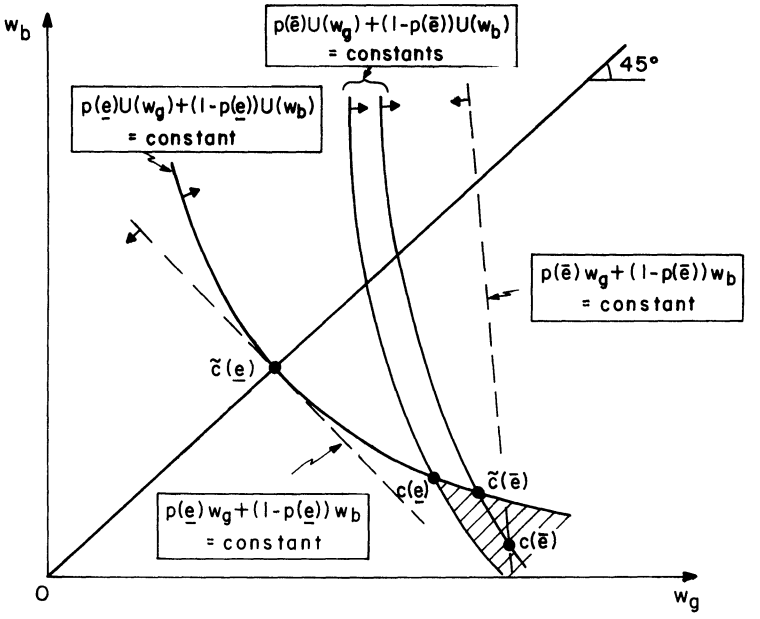

FIGURE 1

with points in the (w_g, w_b) space, and depicts an initial contract c where type $\underline{e}$ does not have full insurance. Note that type $\bar{e}$'s indifference curve is always steeper than type $\underline{e}$'s, as the type with the higher risk of a low outcome values insurance more. The shaded area depicts the set of compensation schemes $c(\bar{e})$ for type $\bar{e}$ that are incentive compatible. Suppose that the initial contract c_1 is the incentive compatible contract $c = (c(\underline{e}), c(\bar{e}))$, and that the principal at the interim stage offers $c_2 = \tilde{c} = (\tilde{c}(\underline{e}), c(\bar{e}))$, where $\tilde{c}(e)$ is the full insurance compensation scheme that yields the same expected utility as $c(\underline{e})$ to type $\underline{e}$. The new contract is incentive compatible, leaves both types indifferent, and lowers the principal's wage bill when the agent has type $\underline{e}$. Hence, only contracts that offer full insurance of type $\underline{e}$ are renegotiation proof, which explains part (a) of Lemma 2.1. To see why part (c) holds, suppose that the initial contract c_1 is $\tilde{c} = (\tilde{c}(\underline{e}), c(\bar{e}))$, so that IIC$(\underline{e})$ is not binding. The principal can offer a bit more insurance to type $\bar{e}$ without violating incentive compatibility. For instance, he can offer contract $c_2 = \tilde{\tilde{c}} = (\tilde{c}(\underline{e}), \tilde{c}(\bar{e}))$, where $\tilde{c}(\bar{e})$ is the intersection of type $\underline{e}$'s indifference curve through $\tilde{c}(\underline{e})$. Type $\underline{e}$ is left indifferent; type $\bar{e}$ is made better off, and the wage bill for type $\bar{e}$ is decreased.

Next, we show that a contract c satisfying (a), (b), and (c) in Lemma 2.1 is renegotiation-proof if and only if x is not too high:

LEMMA 2.2: *Let* $c = \{U_g(e), U_b(e)\}_{e=\underline{e},\bar{e}}$ *be a feasible contract satisfying*

(a) $U_g(\underline{e}) = U_b(\underline{e}) = \underline{U}$;

(b) $U_g(\bar{e}) \geqslant U_b(\bar{e})$;

and

$$\text{(c)} \qquad p(\underline{e})U_g(\underline{e}) + (1-p(\underline{e}))U_b(\underline{e}) = p(\underline{e})U_g(\bar{e}) + (1-p(\underline{e}))U_b(\bar{e}).$$

There exists $x^*(c) \in [0,1]$ *such that* c *is renegotiation proof if and only if* $x \leqslant x^*(c)$. *If* $U_g(\bar{e}) = U_b(\bar{e})$, $x^*(c) = 1$. *If* $U_g(\bar{e}) > U_b(\bar{e})$, $x^*(c)$ *is the unique solution of*:

$$\text{(2.1)} \qquad \frac{x^*(c)}{1-x^*(c)} = \left(\frac{\Phi'(\underline{U})}{\phi'(U_g(\bar{e})) - \Phi'(U_b(\bar{e}))}\right)\left(\frac{p(\bar{e}) - p(\underline{e})}{p(\bar{e})(1-p(\bar{e}))}\right).$$

PROOF: Consider an initial contract $c = (c(\underline{e}), c(\bar{e}))$ satisfying (a), (b), and (c), and consider the minimization of $M(\tilde{c}, x)$ in Definition 2.1. From the proof of Lemma 2.1, we know that we can restrict attention to interim contracts $\tilde{c} = (\tilde{c}(\underline{e}), \tilde{c}(\bar{e}))$ that satisfy (a), (b), and (c). Let $v = \tilde{U}_g(\underline{e}) = \tilde{U}_b(\underline{e})$, and let $\tilde{c}(\bar{e}) = \{\tilde{U}_g^v, \tilde{U}_b^v\}$ denote the compensation scheme defined by IIC($\underline{e}$):

$$p(\underline{e})\tilde{U}_g^v + (1-p(\underline{e}))\tilde{U}_b^v = v$$

and IIR($\bar{e}$):

$$p(\bar{e})\tilde{U}_g^v + (1-p(\bar{e}))\tilde{U}_b^v = p(\bar{e})U_g(\bar{e}) + (1-p(\bar{e}))U_b(\bar{e}).$$

Figure 2 depicts $\tilde{c}(\bar{e})$ in the wage space. For a given choice of $\tilde{c}(\underline{e})$ (or v), $\tilde{c}(\bar{e})$ is the best interim feasible compensation scheme for type $\bar{e}$ from the principal's viewpoint.

Let $M(v)$ be the expected wage bill for contract $\tilde{c}$:

$$\text{(2.2)} \qquad M(v) = x\left[p(\bar{e})\Phi\left(\tilde{U}_g^v\right) + (1-p(\bar{e}))\Phi\left(\tilde{U}_b^v\right)\right] + (1-x)\Phi(v).$$

Differentiating (2.2) yields

$$\text{(2.3)} \qquad \frac{dM}{dv} = -\left(\frac{xp(\bar{e})(1-p(\bar{e}))}{p(\bar{e}) - p(\underline{e})}\right)\left(\Phi'\left(\tilde{U}_g^v\right) - \Phi'\left(\tilde{U}_b^v\right)\right) + (1-x)\Phi'(v),$$

where $\Phi'(U) = d\Phi/dU$ and where we have used IIC($\underline{e}$) and IIR($\bar{e}$) to compute $d\tilde{U}_g^v/dv$ and $d\tilde{U}_b^v/dv$. Since U'' is negative by assumption, $\Phi'' = -U''/(U')^2$ is positive. (2.3) and the fact that $\tilde{U}_g^v \geqslant \tilde{U}_b^v$ imply that $d^2M/dv^2 < 0$. Therefore contract c is renegotiation-proof for distribution x if and only if $dM/dv \geqslant 0$ at $v = \underline{U}$, which from (2.3) amounts to $x \leqslant x^*(c)$. *Q.E.D.*

Again, the idea of Lemma 2.2 can be grasped from Figure 2. Suppose that from the initial contract c, the principal offers to raise $\underline{U}$ to $(\underline{U} + \delta\underline{U})$ (i.e., move from $c(\underline{e})$ to $\tilde{c}(\underline{e})$). This increases the wage bill by $(1-x)\Phi'(\underline{U})\delta\underline{U}$, but allows the principal to give more insurance to type $\bar{e}$ (i.e., move from $c(\bar{e})$ to $\tilde{c}(\bar{e})$). Let $(\delta U_g < 0, \delta U_b > 0)$ denote a change in type $\bar{e}$'s compensation scheme that keeps type $\bar{e}$ on its indifference curve: $p(\bar{e})\delta U_g + (1-p(\bar{e}))\delta U_b = 0$.

This makes $\tilde{c}(\bar{e})$ more attractive to type $\underline{e}$ as well, which is why it was necessary to increase $\underline{U}$:

$$\delta\underline{U} = p(\underline{e})\delta U_g + (1-p(\underline{e}))\delta U_b.$$

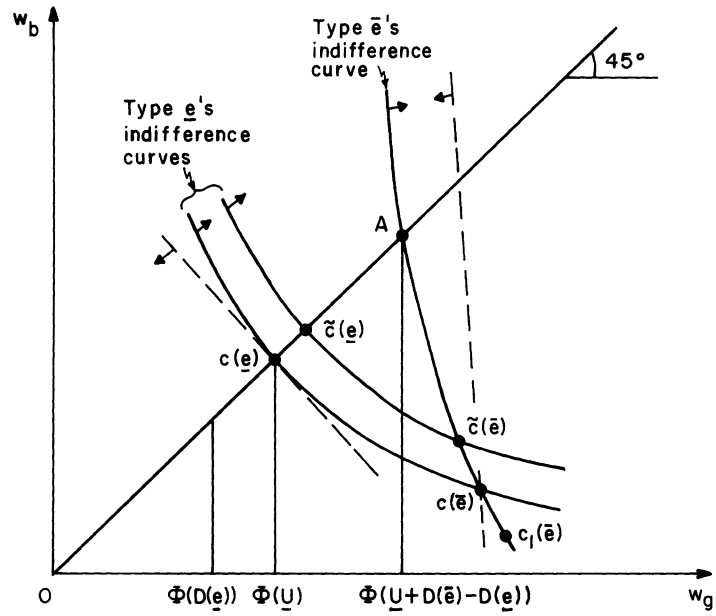

FIGURE 2

The expected gain in the wage bill due to better insurance of type $\bar{e}$ is then $x[\,p(\bar{e})\Phi'(U_g(\bar{e}))\delta U_g + (1-p(\bar{e}))\Phi'(U_b(\bar{e}))\delta U_b]$. For the contract c to be renegotiation-proof, it must be the case that the expected gain from insuring type $\bar{e}$ be lower than the expected increase in type $\underline{e}$'s wage, which yields Lemma 2.2.

3. EX-ANTE CONTRACTS AND EQUILIBRIUM

3A. *Renegotiation-proof Ex-ante Contracts*

We now analyze the agent's choice of effort and the *ex-ante* choice of contract. We already noted that any equilibrium outcome of the game described in A is also the outcome of a continuation equilibrium for a contract c_1 that is renegotiation-proof for the induced distribution x. This renegotiation-proofness principle considerably simplifies the quest for an upper bound on the principal's equilibrium payoff. After obtaining such a bound, we then show that it is attained in some equilibrium. Section 3B shows that there is an initial contract with a unique continuation equilibrium that attains this bound.

We noted previously that, in equilibrium, the probability x that the agent chooses the high effort must be strictly less than one. (Otherwise, renegotiation-proofness would imply full insurance, and foreseeing this the agent would choose the low effort with probability one.) In contrast, inducing $x = 0$ is simple: it suffices to give a single compensation scheme $c(\underline{e}) = c(\bar{e}) = \{\underline{U}, \underline{U}\}$. Because a full-insurance contract is always renegotiation-proof, it induces a choice of effort $\underline{e}$, and the cost minimizing choice of $\underline{U}$ to induce $x = 0$ is $\Phi(D(\underline{e}))$.

Let us now focus on renegotiation-proof contracts that are consistent with mixing by the agent:

DEFINITION 3.1: A contract c is *consistent with distribution* $x \in (0,1)$ *and rent* $R \geqslant 0$ if (i) it is renegotiation-proof for distribution x (i.e., satisfies condition (a), (b), (c) of Lemma 2.2 and $x \leqslant x^*(c)$), and (ii) it satisfies *ex-ante* incentive compatibility (AIC):

$$\begin{aligned} & p(\bar{e})U_g(\bar{e}) + (1-p(\bar{e}))U_b(\bar{e}) - D(\bar{e}) \\ & \quad = p(\underline{e})U_g(\underline{e}) + (1-p(\underline{e}))U_b(\underline{e}) - D(\underline{e}) \\ & \quad = R. \end{aligned}$$

In order to randomize between the two effort levels, the agent must obtain the same utility for both, hence AIC. Since $D(\bar{e}) > D(\underline{e})$, AIC implies that type $\bar{e}$ has a higher interim utility than type $\underline{e}$. Thus AIC rules out the first class of contracts in Lemma 2.1. We will say that "the agent enjoys a rent" if $R > 0$. Note that *ex-ante* incentive compatibility and interim incentive compatibility imply that the agent does not benefit from choosing some effort e and announcing effort $\tilde{e} \neq e$.

Suppose that the principal offers a contract c consistent with distribution $x \in (0,1)$ and rent $R \geqslant 0$. A continuation equilibrium is an equilibrium of the subgame in which contract c has been chosen. We claim that there exists a continuation equilibrium in which the agent accepts the contract, chooses $\bar{e}$ with probability x, and the contract is not renegotiated: If the principal expects probability x, he cannot gain from renegotiating, so no renegotiation is an equilibrium at the interim stage. Anticipating no renegotiation, the agent is indifferent between the two efforts from AIC and is willing to choose $\bar{e}$ with probability x. Last, the agent is willing to accept c as $R \geqslant 0$.

To summarize, the upper bound on the principal's payoff in any equilibrium is equal to the principal's highest payoff among contracts c that are consistent with distribution $x \in (0,1)$ and rent $R \geqslant 0$, or else is equal to $[I(\underline{e}) - \Phi(D(\underline{e}))]$.

We now show that for each rent R the specification of the distribution and the rent fully determines the renegotiation-proof contract. To see why, note that AIC and IIC($\underline{e}$) yield three linear equations in three unknowns $\{U_g(\bar{e}), U_b(\bar{e}), R\}$ which have a single solution. (The intuition can again be grasped from Figure 2: R determines $\underline{U} = D(\underline{e}) + R$ and hence $c(\underline{e})$. Then $c(\bar{e})$ must belong to type $\underline{e}$'s indifference curve through $c(\underline{e})$ and furthermore satisfy *ex-ante* indifference between the two actions. That is, the certainty equivalent wage corresponding to $c(\bar{e})$ and probabilities $(p(\bar{e}), 1-p(\bar{e}))$, which is depicted by A in Figure 2, must equal $\Phi(\underline{U} + D(\bar{e}) - D(\underline{e}))$.)

Let $U^R(\underline{e}) = D(\underline{e}) + R$, and let $(U_g^R(\bar{e}), U_b^R(\bar{e}))$ be the solution to IIC($\underline{e}$) and AIC. From the linearity of the system, we have: $U^R(\underline{e}) = U^0(\underline{e}) + R, U_g^R(\bar{e}) =$

$U_g^0(\bar{e}) + R$, and $U_b^R(\bar{e}) = U_b^0(\bar{e}) + R$, where

$$(3.1)\qquad U_g^0(\bar{e}) = \left[(1-p(\underline{e}))D(\bar{e}) - (1-p(\bar{e}))D(\underline{e})\right]/\left[p(\bar{e}) - p(\underline{e})\right],$$
$$U_b^0(\bar{e}) = \left[p(\bar{e})D(\underline{e}) - p(\underline{e})D(\bar{e})\right]/\left[p(\bar{e}) - p(\underline{e})\right],$$
$$U^0(\underline{e}) = D(\underline{e}).$$

LEMMA 3.1: *If there exists a contract consistent with distribution x, $x > 0$, and rent R, it is the unique solution of*

(i) $\quad c(\underline{e}) = c^R(\underline{e}) = \{U^0(\underline{e}) + R, U^0(\underline{e}) + R\}$ *and*

(ii) $\quad c(\bar{e}) = c^R(\bar{e}) = \{U_g^0(\bar{e}) + R, U_b^0(\bar{e}) + R\}$.

Conversely, the contract $c^R = \{c^R(\underline{e}), c^R(\bar{e})\}$ is consistent with distribution $x > 0$ and rent R if and only if $x \leqslant x^*(c^R) \equiv x^*(R)$ (by a slight abuse of notation).

PROPOSITION 3.1: *The principal's payoff is at most the maximum of three numbers:*

0 (*obtained by not offering a contract*);

$I(\underline{e}) - \Phi(D(\underline{e}))$ (*obtained by offering the full insurance compensation scheme at wage $\Phi(D(\underline{e}))$; and*:

$$(3.2)\qquad W = \sup_{R \geqslant 0} \Big\{ x^*(R)\left[I(\bar{e}) - p(\bar{e})\Phi\left(U_g^0(\bar{e}) + R\right) - (1-p(\bar{e}))\Phi\left(U_b^0(\bar{e}) + R\right)\right] + (1-x^*(R))\left[I(\underline{e}) - \Phi\left(U^0(\underline{e}) + R\right)\right]\Big\}$$

(*obtained by inducing the agent to randomize*).

Two comments on Proposition 3.1. First, the supremum in (3.2) can be taken to be a maximum: Because the principal's payoff tends to $-\infty$ when R tends to $+\infty$, we can restrict the rents to some compact set $[0, \bar{R}]$, which implies that the supremum is attained. Second, again in (3.2), we assumed that the agent chooses the highest mixing probability x that makes contract c^R consistent with x and R: $x = x^*(R)$. The reason for this is that the principal's payoff given c^R is linear in x, so its maximum over $x \in [0, x^*(R)]$ is either at $x = 0$ or at $x = x^*(R)$. But for $x = 0$, the payoff is equal to $I(\underline{e}) - \Phi(U^0(\underline{e}) + R) \leqslant I(\underline{e}) - \Phi(D(\underline{e}))$ from (3.1).

Now we wish to confirm that a perfect Bayesian equilibrium exists that attains the upper bound of Proposition 3.1:

PROPOSITION 3.2: *There exists a perfect Bayesian equilibrium of the game in which the principal's payoff is equal to the upper bound characterized in Proposition* 3.1.

PROOF: Construct the following equilibrium path: If 0 is the highest number in Proposition 3.1, the principal quits. If not, and $I(\underline{e}) - \Phi(D(\underline{e}))$ is the highest

number, the principal offers the full insurance compensation scheme at wage $\Phi(D(\underline{e}))$. If the third number (W) is strictly greater than the other two, let the principal offer c^R where R is the arg max in (3.2) (which exists as we just noted).

Suppose the principal deviates and offers another contract c. We already know that if a continuation equilibrium exists, it yields the principal at most the equilibrium payoff. Hence, it suffices to prove that a continuation equilibrium exists for any $c = \{U_g(e), U_b(e)\}_{e=\underline{e},\bar{e}}$. To prove this is routine. Fix x in $[0,1]$. To this x, associate the unique solution $\tilde{c}(x) = \{\tilde{U}_g(e), \tilde{U}_b(e)\}(x)$ to the minimization in Definition (2.1) where the initial contract $\{U_g(e), U_b(e)\}$ equals the given contract c. Because $\Phi'' > 0$, $M(\tilde{c}, x)$ is convex in $\{\tilde{U}_g, \tilde{U}_b\}$ and the constraints (i) and (ii) are linear. Thus, the solution to the minimization varies continuously in x. Then consider the correspondence which maps x to the set of distributions y that maximize

$$\Big\{y\Big(p(\bar{e})\tilde{U}_g(\bar{e}) + (1-p(\bar{e}))\tilde{U}_b(\bar{e}) - D(\bar{e})\Big)$$

$$+(1-y)\Big(p(\underline{e})\tilde{U}_g(\underline{e}) + (1-p(\underline{e}))\tilde{U}_b(\underline{e}) - D(\underline{e})\Big)\Big\}.$$

This correspondence is nonempty, upper-hemi-continuous and convex valued. Kakutani's theorem implies that it has a fixed point, which by construction defines a continuation equilibrium when contract c is accepted. Last contract c is accepted if and only if the agent gets at least his reservation utility in the continuation equilibrium. *Q.E.D.*

The interesting case is of course when W is the highest of the three numbers in Proposition 3, which we will assume from now on. For an R that yields W, we will say that c^R is the "optimal renegotiation-proof contract." For the moment, we assume that if the principal offers c^R, the agent accepts and randomizes in the way ($x^*(R)$) that is optimal for the principal. Note, however, that this assumption is heroic at this stage because the agent's indifference might lead to a lower probability x in the continuation equilibrium. In Section 3B, we show that the principal need not worry about the agent's indifference, as she can guarantee W as her unique payoff by offering a different contract.

Assuming that W is the equilibrium payoff, it remains to determine the optimal rent R. To this purpose we investigate how $x^*(R)$ varies with R. Inspecting the right-hand side of (2.1), we see that R influences $x^*(R)$ through its effect on the ratio

$$Q(R) = \Phi'\big(U^0(\underline{e}) + R\big)/\Big[\Phi'\big(U_g^0(\bar{e}) + R\big) - \Phi'\big(U_b^0(\bar{e}) + R\big)\Big],$$

which compares the "demand for insurance" $[\Phi'(U_g^0(\bar{e}) + R) - \Phi'(U_b^0(\bar{e}) + R)]$ to the cost of increasing $\underline{e}$'s utility. The sign of this effect depends on how the ratio Φ''/Φ' varies with U:

LEMMA 3.2: (i) *The sign of* dx^*/dR *is the negative of the sign of*

$$\frac{d}{dU}\left(\frac{\Phi''(U)}{\Phi'(U)}\right).$$

(ii) *If* $U(w)$ *has constant or increasing absolute risk aversion,* $x^*(R)$ *is decreasing in* R; *if* $U(w)$ *has constant relative risk aversion* α, *then the sign of* dx^*/dR *equals* $1-\alpha$.

PROOF: (i) Since $x/(1-x)$ is monotone increasing in x for $x \in (0,1)$, $\operatorname{sign}(dx^*/dR) = \operatorname{sign}(dQ/dR)$. We compute

$$\begin{aligned}(3.3)\qquad \frac{dQ}{dR} \propto\ & \Phi'\big(U_g^0(\bar{e})+R\big)\Phi'\big(U^0(\underline{e})+R\big) \\ & \times\left[\frac{\Phi''\big(U^0(\underline{e})+R\big)}{\Phi'\big(U^0(\underline{e})+R\big)} - \frac{\Phi''\big(U_g^0(\bar{e})+R\big)}{\Phi'\big(U_g^0(\bar{e})+R\big)}\right] \\ & + \Phi'\big(U^0(\underline{e})+R\big)\Phi'\big(U_b^0(\bar{e})+R\big) \\ & \times\left[\frac{\Phi''\big(U_b^0(\bar{e})+R\big)}{\Phi'\big(U_b^0(\bar{e})+R\big)} - \frac{\Phi''\big(U^0(\underline{e})+R\big)}{\Phi'\big(U^0(\underline{e})+R\big)}\right],\end{aligned}$$

where $\propto$ means "proportional to." Since $U'' < 0$,

$$\Phi'' = \frac{-U''(\Phi')^2}{U'} > 0.$$

Since $U_g^0(\bar{e}) > U^0(\underline{e}) > U_b^0(\bar{e})$, the sign of dQ/dR is positive, zero, or negative as Φ''/Φ' is decreasing, constant, or increasing in U. This proves (i).

(ii) The coefficient of absolute risk aversion, $-U''/U'$ is equal to $\Phi''/(\Phi')^2$. Since Φ'' is positive, Φ' is an increasing function, and so Φ''/Φ' is increasing if $\Phi''/(\Phi')^2$ is constant or increasing. Next, for constant relative risk aversion α, $-wU''/U' = \alpha = \Phi''\Phi/(\Phi')^2$. Thus $\Phi''/\Phi' = \alpha\Phi'/\Phi$, and

$$\begin{aligned} d(\Phi''/\Phi')/dU &= \alpha\big[\big(\Phi''\Phi - (\Phi')^2\big)/\Phi^2\big] \\ &= [\alpha\Phi'/\Phi][\Phi''/\Phi' - \Phi'/\Phi] \\ &= \alpha(\Phi'/\Phi)^2(\alpha-1). \end{aligned}$$

Q.E.D.

The intuition for Lemma 3.2 is that dx^*/dR is positive if increasing the agent's rent reduces type $\bar{e}$'s demand for insurance faster than it increases the cost $\Phi'(U^0(\underline{e})+R)$ of increasing $\underline{e}$'s utility. With increasing absolute risk aversion, increasing the rent *increases* the demand for insurance, and thus does not relax the renegotiation-proofness constraint. For the opposite conclusion to obtain, the agent's risk aversion must decline sufficiently quickly in his wealth. Note that with constant relative risk aversion α, the agent's absolute risk aversion is α/w, which decreases more quickly when α is larger. Thus we

should expect that $dx^*/dR > 0$ if the agent's relative risk aversion is sufficiently high.

If $x^*(R)$ is decreasing or constant in R, the principal will never choose $R > 0$: Choosing a larger R increases compensation costs and will not improve the distribution of effort levels. However, if $x^*(R)$ is increasing, and the difference $I(\bar{e}) - I(\underline{e})$ is sufficiently large, it will be optimal to choose $R > 0$: Here the increased compensation costs are outweighed by the fact that with a higher rent a better distribution of efforts becomes renegotiation-proof.

PROPOSITION 3.3: (*i*) *If Φ''/Φ' is monotone decreasing or constant, the optimal renegotiation-proof contract has $R = 0$.*

(*ii*) *If Φ''/Φ' is increasing, there is a critical number Δ, $0 < \Delta < \infty$, such that the form of the optimal renegotiation-proof contract depends on the sign of $I(\bar{e}) - I(\underline{e}) - \Delta$. If this expression is negative, then the optimal contract is the same as in case (i) above. If it is positive, the optimal contract has $R > 0$: The principal gives the agent an ex-ante rent in order to relax the renegotiation-proofness constraint.*

PROOF: In case (i), it is clear that the principal should choose $R = 0$ from (3.2). In case (ii), note that the increase in expected revenue that the principal obtains by giving the agent a rent $R > 0$ is $(x^*(R) - x^*(0))(I(\bar{e}) - I(\underline{e}))$ which is increasing in $(I(\bar{e}) - I(\underline{e}))$, so there is a Δ such that the principal gains by offering a nonzero rent if and only if $(I(\bar{e}) - I(\underline{e})) > \Delta$. *Q.E.D.*

3B. *Uniqueness of the Continuation Equilibrium*

Section 3A derived the upper bound on the principal's payoff in the set of Nash equilibria and showed that there exists a perfect Bayesian equilibrium in which this upper bound is obtained. We now show that the overall game has a unique equilibrium payoff by exhibiting contracts c_1 such that each c_1 has the same, unique, continuation equilibrium, and the continuation equilibrium yields the upper bound to the principal's payoff.

These initial contracts are not renegotiation-proof. Each of them offers the same scheme to both types, and is renegotiated to the menu $\{c(\underline{e}), c(\bar{e})\}$ of compensation schemes characterized in Proposition 3.2 (i.e., satisfying (a) and (c) of Lemma 2.1, AIC, (2.1), and (3.2)). While these initial type-invariant contracts guarantee the principal the upper bound, the optimal renegotiation-proof contract does not. If the principal offers this renegotiation-proof contract, any $x \leqslant x^*$ followed by the absence of renegotiation at the interim stage is a continuation equilibrium. In the same way that the revelation principle allows us to restrict attention to truthful revelation games to derive all equilibrium payoffs of a game but does not guarantee uniqueness of equilibrium (i.e., truthful revelation) in the revelation games, the renegotiation proofness principle allows us to focus on renegotiation-proof contracts to obtain the upper bound on the principal's equilibrium payoff but does not ensure uniqueness in

the continuation game following the acceptance of the contract.

Subsection 3C uses an argument similar to Harsanyi (1973) to show that if the agent *ex-ante* has private information about his preferences, the upper bound on the principal's payoff can be obtained as the unique continuation equilibrium payoff associated with a renegotiation-proof contract; that as the principal's uncertainty about this *ex-ante* private information becomes small the upper bound converges to that derived in Proposition 3.1; and last that the agent plays a pure strategy. Thus uniqueness of the continuation equilibrium is consistent with renegotiation-proofness.

PROPOSITION 3.4: *In any perfect Bayesian equilibrium (PBE), the payoffs are the optimal values described in Proposition* 3.2.

REMARK: The restriction to PBE in this proposition is important, as the principal's payoff can be lower in some Nash equilibria. Consider, for example, the strategies: "If the initial contract offer is the optimal renegotiation-proof menu, the agent accepts and sets $x = x^*/2$. The agent refuses all other initial contracts."

PROOF: Let $c^*(\underline{e}), c^*(\bar{e}), R^*, x^* = x^*(R^*)$ the optimal compensation schemes, rent, and mixing probability in the optimal outcome satisfying (a) and (c) of Lemma 2.1, AIC, (2.1), and (3.2). Consider *any* compensation scheme $c_1(\bar{e}) = \{U_g^1(\bar{e}), U_b^1(\bar{e})\}$ that lies on type $\bar{e}$'s indifference curve through $c^*(\bar{e})$ and is strictly more risky than $c^*(\bar{e})$ (see Figure 2, where $c(\cdot)$ is now the optimal contract $c^*(\cdot)$):[4]

$$p(\bar{e})U_g^1(\bar{e}) + (1-p(\bar{e}))U_b^1(\bar{e}) = p(\bar{e})U_g^*(\bar{e}) + (1-p(\bar{e}))U_b^*(\bar{e}) \tag{3.4}$$

and

$$U_b^1(\bar{e}) < U_b^*(\bar{e}). \tag{3.5}$$

Under these conditions, the agent receives an *ex-ante* rent of R^* if he chooses effort $\bar{e}$ and the contract $c_1(\bar{e})$ is not renegotiated. Since the agent receives rent R^* from choosing $\underline{e}$ under contract $c^*(\bar{e})$, the single crossing condition implies that $c_1(\bar{e})$ gives rent less than R^* to effort $\underline{e}$. If $x = 0$ the principal would renegotiate to a riskless contract that gives type $\underline{e}$ the same rent as the initial contract, and since this initial rent is less than R^* the agent would do strictly better to choose $\bar{e}$. Hence in equilibrium x must be greater than zero.

Suppose the principal offers an initial contract composed of a single compensation scheme $c_1 = \{c_1(\bar{e}), c_1(\bar{e})\}$ and the agent accepts. We know a continuation equilibrium exists, with some probability x of high effort, new contract offer c_2, acceptance/rejection of c_2 and announcement of type by agent. We claim that

[4] This is the only place in the paper in which we need the assumption that utilities are unbounded below. What we actually need is that $U_b(\bar{e})$ be strictly greater than the lower bound on utilities, in order for such a feasible contract $c_1(\bar{e})$ to exist.

$x = x^*, c_2 = \{c^*(\underline{e}), c^*(\bar{e})\}$, c_2 is accepted and the agent announces his true type. For any x, the principal minimizes $M(\tilde{c}, x)$ as in Definition 2.1. From Lemma 2.1, the solution $c_2 = \{\tilde{c}(\underline{e}), \tilde{c}(\bar{e})\}$ gives full insurance to type $\underline{e}$, and satisfies IIR($\bar{e}$) and IIC($\underline{e}$):

$$\tilde{U}_g(\underline{e}) = \tilde{U}_b(\underline{e}) \equiv \underline{\tilde{U}}, \tag{3.6}$$

$$p(\bar{e})\tilde{U}_g(\bar{e}) + (1-p(\bar{e}))\tilde{U}_b(\bar{e}) = p(\bar{e})U_g^1(\bar{e}) + (1-p(\bar{e}))U_b^1(\bar{e}), \tag{3.7}$$

$$\underline{\tilde{U}} = p(\underline{e})\tilde{U}_g(\bar{e}) + (1-p(\underline{e}))\tilde{U}_b(\bar{e}). \tag{3.8}$$

Let $\underline{U} = p(\underline{e})U_g(\underline{e}) + (1-p(\underline{e}))U_b(\underline{e})$. From Lemma 2.2:

$$\underline{\tilde{U}} \lesseqgtr \underline{U} \quad \text{if} \quad x \lesseqgtr x^*. \tag{3.9}$$

In particular $\{\tilde{c}(\underline{e}), \tilde{c}(\bar{e})\}$ coincides with $\{c^*(\underline{e}), c^*(\bar{e})\}$ if and only if $x = x^*$. Now consider the choice of x: If $x < x^*$, (3.9), (3.4), (3.7), and AIC imply

$$\begin{aligned} \underline{\tilde{U}} - D(\underline{e}) < \underline{U} - D(\underline{e}) &= p(\bar{e})U_g^1(\bar{e}) + (1-p(\bar{e}))U_b^1(\bar{e}) - D(\bar{e}) \\ &= p(\bar{e})\tilde{U}_g(\bar{e}) + (1-p(\bar{e}))\tilde{U}_b(\bar{e}) - D(\bar{e}). \end{aligned} \tag{3.10}$$

Hence the agent strictly prefers the high effort ($x = 1$), a contradiction. Similarly, if $x > x^*$, (3.9), (3.4), (3.7), and AIC imply

$$\underline{\tilde{U}} - D(\underline{e}) > p(\bar{e})\tilde{U}_g(\bar{e}) + (1-p(\bar{e}))\tilde{U}_b(\bar{e}) - D(\bar{e}) \tag{3.11}$$

so that the agent strictly prefers the low effort ($x = 0$), a contradiction. We conclude that for contract c_1, $x = x^*$, and $\{\tilde{c}(\underline{e}), \tilde{c}(\bar{e})\} = \{c^*(\underline{e}), c^*(\bar{e})\}$ is the unique continuation equilibrium outcome.[5] *Q.E.D.*

Thus the principal can guarantee himself the payoff of Proposition 3.2 in any PBE. Since this is the highest payoff in any Nash equilibrium, every PBE must give the principal the optimal payoff.

3C. *Renegotiation-proofness and Purification*

Although the optimal payoff cannot be guaranteed with a renegotiation-proof contract, payoffs close to the optimal payoff can be ensured with renegotiation-proof contracts when the model is enriched to allow the agent to have a "small" amount of private information about his preferences.

This "enrichment" is simply that of Harsanyi's (1973) defense of mixed strategies. The idea is to introduce a little bit of uncertainty about the difference

[5] Note that the agent is (1) *ex-ante* indifferent between accepting and rejecting c_1 if $R^* = 0$, (2) indifferent between efforts $\underline{e}$ and $\bar{e}$, (3) indifferent between announcing $\underline{e}$ or $\bar{e}$ after having exerted effort $\underline{e}$. As in other incentive problems, the first and the third indifferences can be broken by giving a rent $\varepsilon > 0$ and relaxing IIC($\underline{e}$) slightly, respectively, so that the agent's resolving the first indifference in favor of acceptance and the third in favor of truthtelling is the only behavior consistent with equilibrium. We saw that the second indifference and the agent's choosing probability x^* is a necessary feature of equilibrium behavior.

$D(\bar{e}) - D(\underline{e})$, which defines an *ex-ante* type for the agent. Fixing the optimal contract derived above, types with a high differential strictly prefer the low effort while types with a low differential strictly prefer the high effort. The equilibrium is then in pure strategies, but is close to our mixed strategy equilibrium. We will present a proof that this purification argument works for the two-action case, but we believe that versions of the result should obtain with more effort levels, and also in other models where the principal wishes to induce a mixed-strategy response, such as Laffont-Tirole (1988).

We start from the model of Section 2, and continue to assume that W is the highest number in Proposition 3.1. We consider a family of "elaborations" of the original model, indexed by n. In each elaboration, the situation is as described in Section 2, with the following exception: The disutility of the high effort is $D(\bar{e}) + \varepsilon$, where ε is private information to the agent before contracting. The private information ε is distributed according to the prior distribution $G^n(\varepsilon)$ which is common knowledge. These elaborations are a small change to the original game in the sense that the supports of the family of distribution G^n converge to the point zero. That is, $\lim_{n\to\infty}\{G^n(\varepsilon) - G^n(-\varepsilon)\} = 1$ for all ε. We assume that each G^n has convex support and is absolutely continuous. It is easy to see that introducing the private information ε would make only a small difference in the principal's expected payoff if he knew ε before contracting. That is, $\lim_{n\to\infty} \overline{W}^n = W$, where W is the principal's maximized payoff as determined in Section 2, and $\overline{W}^n$ is the expected value of the corresponding maximum when the principal knows ε and ε has distribution G^n.

Let $\tilde{W}^n$ be the principal's maximized payoff in G^n when ε is private information. Let $\tilde{x}^n$ be the marginal probability over all types ε that the agent chooses effort $\bar{e}$.

PROPOSITION 3.5: *$\tilde{W}^n \to W$, and $\tilde{W}^n$ can be attained with a renegotiation-proof contract $\tilde{c}^n$ such that for G^n-almost all ε the agent has a unique optimal action, and such that $\tilde{x}^n$ converges to the optimal level $x^*(R)$ obtained in Proposition* 3.1. *Furthermore, the continuation equilibrium following acceptance of $\tilde{c}^n$ is unique.*

REMARK: Because the contracts we will construct make almost all types have a unique optimal choice of effort, mixed strategies are irrelevant and the issue of how the agent behaves when indifferent is moot.

PROOF: See Fudenberg-Tirole (1988).

As we mentioned, our argument is closely related to Harsanyi's. There are two minor differences. The first is technical: we are unaware of purification results which cover extensive form games (purification theorems are usually on the normal form), or the type of continuum of actions considered here. Second, the Harsanyi argument shows how a mixed strategy equilibrium has a pure strategy neighbor. Our conclusion here is slightly stronger: the purification yields a unique equilibrium payoff and a unique continuation equilibrium.

4. LINK WITH THE COMMITMENT CASE

One way to assess the implications of renegotiation in our model is to compare the optimal renegotiation-proof contract to the optimal contract when renegotiation is not possible. We have seen that renegotiation constraints may lead the principal to induce a different distribution over effort levels than he would choose under commitment. We show that this is sometimes the *only* way in which they differ. More precisely, we show that a contract is consistent with a given distribution x and rent R (in the sense of Definition 3.1) only if it is the cost-minimizing way to induce distribution x with rent R under the *ex-ante* incentive constraints alone. Since the optimal rent under commitment is zero, whenever the optimal rent under renegotiation is zero the force of renegotiation is only to change *which* distribution is implemented and *not* to change the implementation of a given distribution. This follows from the special nature of our model, and is not true in more general settings, as we explain in our concluding remarks. However, the result below does apply to more general utility functions, action spaces, and outcome spaces; to make this clear we now generalize our notation.

Let E denote the set of feasible efforts e, Y the set of possible outcomes y. The agent has a von Neumann-Morgenstern utility function $V(w, e)$. The effort e induces a cumulative distribution function $H(y; e)$, so the principal's expected income is

$$I(e) = \int y\, dH(y; e) = E_e y.$$

Suppose that the principal can commit not to renegotiate the original contract and that the agent's reservation utility is R. Following Grossman and Hart (1983) we find it useful to consider the cost-minimizing compensation scheme for a given level of effort.

For a fixed effort e, this problem is to choose $w(y; e)$ to minimize $E_e w(y; e)$ subject to the constraints

$$(IR) \qquad E_e V(w(y; e), e) \geqslant R$$

and

$$(IC) \qquad L(e, e) \equiv E_e V(w(y; e), e) \geqslant E_{e'} V(w(y; e), e') \equiv L(e', e) \quad \text{for all } e'.$$

Here $L(e', e)$ is the agent's utility if he chooses effort e' while the principal offers the contract corresponding to e. Let $w_c^R(y; e)$ denote the cost minimizing compensation scheme with rent R for effort e where "c" stands for "commitment." (We will assume that this scheme is unique. If it is not, replace "the commitment solution" by "a commitment solution" in the statement of Proposition 4.1.)

PROPOSITION 4.1: *If a contract $c = \{w(y; e)\}$ is consistent with a distribution F over effort levels with support $E^* \subseteq E$ and rent R, then it is the commitment solution for each equilibrium level of effort: for F-almost every $e \in E^*$, $w(y; e) \equiv w_c^R(y; e)$.*

PROOF: Let R denote the equilibrium utility of the agent. Given the equilibrium distribution F with support E^*, the principal can offer second-period contract $\{\tilde{w}(y;e)\}_{e\in E^*} = \{w_c^R(y;e)\}_{e\in E^*}$. If this menu is interim incentive compatible, the agent's utility is unaffected. The principal's expected wage bill decreases for each $e \in E^*$ because $w_c^R(\cdot;e)$ is by definition the cost-minimizing compensation scheme to induce e; it strictly decreases for a set of positive measure unless $w(\cdot;\cdot)$ differs from $w_c^R(\cdot;\cdot)$ only on a set of measure zero. Now we claim that (*IC*) *implies interim incentive compatibility*. For, assume that there exists an effort e and an announcement $\tilde{e}$ in E^* such that $L(e,\tilde{e}) > L(e,e)$. Because $L(e,e) = L(\tilde{e},\tilde{e})$ for all $(e,\tilde{e})$ in $(E^*\times E^*)$ (from optimality of the effort choice for the agent), we have $L(e,\tilde{e}) > L(\tilde{e},\tilde{e})$, which implies that the commitment scheme $w_c^R(\cdot;\tilde{e})$ (meant to induce $\tilde{e}$) is not incentive compatible, a contradiction.[6] *Q.E.D.*

5. CONTINUUM MODEL

We now extend our results to a continuum of efforts and two outcomes. Appendix 2 shows how to extend them to continua of efforts and outcomes.

5A. *The Optimal Renegotiation-proof Contract*

We now generalize our earlier results to a continuum of effort levels. The agent chooses effort $e \in E = [\underline{e}, +\infty)$, where the lower bound $\underline{e}$ should be interpreted as the lowest effort level that cannot be directly detected by the principal. The agent has (*ex-ante*) separable utility

$$U(e) \equiv p(e)U_g(e) + (1-p(e))U_b(e) - D(e),$$

where the probability $p(e)$ of a good outcome belongs to $(0,1)$, is increasing and strictly concave in effort ($\dot{p}(e) > 0, \ddot{p}(e) < 0$), and the disutility of effort is increasing and strictly convex in effort ($\dot{D}(e) \geqslant 0, \ddot{D}(e) > 0$), so that $\dot{D}$ is strictly positive except perhaps at $\underline{e}$.

We will see that as in the discrete case, every renegotiation-proof contract either induces $e = \underline{e}$ with probability one or induces the agent to randomize over effort levels. A strategy for the agent is a cumulative distribution function over effort levels, i.e., an increasing right continuous function $F(e)$ taking values in $[0,1]$. Let E^* denote the support of F.

Given the renegotiation constraints, the principal will wish the agent to play a mixed strategy, so we once again consider menus of contracts indexed by $e, c(e) = \{U_g(e), U_b(e)\}$. Fixing an *ex-ante* contract, we let $V(e) = p(e)U_g(e) +$

[6] Let us note that with nonseparable utility functions, the optimal contacts under either commitment or renegotiation can differ from those we have derived because different constraints may be binding. For example, it might be that the agent's risk aversion to income gambles is increasing in his effort, in which case providing $\underline{e}$ with full insurance at the interim stage might violate $\tilde{e}$'s interim IC constraint. (Note that we have restricted attention to deterministic contracts. Under nonseparability, stochastic contracts may be optimal. But it is easy to see that such contracts cannot differ much from deterministic contracts for small nonseparability.)

$(1-p(e))U_b(e)$ denote the agent's interim utility when he has chosen contract $c(e)$ and effort level e.

DEFINITION 5.1: A contract menu is *ex-ante incentive compatible with rent R for support* E^* if all efforts e in E^* yield the same *ex-ante* rent R, and no choice of effort yields a higher rent.

The following lemma shows that the incentive constraints and the rent R completely determine the form of these single contracts except at $\underline{e}$.

LEMMA 5.1: *A menu* $\{U_g^R(e), U_b^R(e)\}_{e\in E^*}$ *is ex-ante incentive compatible with rent R for support* E^* *iff* $U_g^R(e) = U_g^0(e) + R$, *and* $U_b^R(e) = U_b^0(e) + R$ *for all* $e \in E^*$, *where* $\{U_g^0(e), U_b^0(e)\}$ *is the solution (unique on* $E^* - \{\underline{e}\}$*) of*

$$p(e)U_g^0(e) + (1-p(e))U_b^0(e) - D(e) = 0 \quad \forall e \in e^*, \tag{5.1a}$$

$$\dot{p}(e)\left(U_g^0(e) - U_b^0(e)\right) - \dot{D}(e) = 0 \quad \forall e \in E^* - \{\underline{e}\}, \tag{5.1b}$$

$$\dot{p}(\underline{e})\left(U_g^0(\underline{e}) - U_b^0(\underline{e})\right) - \dot{D}(\underline{e}) \leqslant 0 \qquad \text{if } \underline{e} \in E^*,$$

$$p(\tilde{e})U_g^0(e) + (1-p(\tilde{e}))U_b^0(e) - D(e) \leqslant 0 \quad \forall e \in E^*, \forall \tilde{e} \in E. \tag{5.1c}$$

PROOF: First, we claim any *ex-ante* incentive compatible contract for E^* with rent 0 must satisfy system (5.1). Equation (5.1a) simply says that the agent obtains zero rent by choosing an effort in E^* and the corresponding contract; (5.1c) says no choice of effort yields higher rent. (5.1b) is the local version of the incentive compatibility constraint. If it is not satisfied at $\tilde{e}$, the agent can do better by choosing contract $c(\tilde{e})$ and an effort level that is slightly different. Thus, conditions (5.1) are necessary. For $e \neq \underline{e}$, equations (5.1a) and (5.1b) are a nonsingular linear system of two equations in two unknowns, and thus have a unique solution $\{U_g^0(e), U_b^0(e)\}$. This solution has the property that $U_g^0(e) > U_b^0(e)$ for all $e > \underline{e}$. Since $\ddot{p} < 0$ and $\ddot{D} > 0$, this implies that the agent's *ex-ante* payoff is strictly concave in his effort, so that the solution to (5.1a) and (5.1b) also satisfies (5.1c). Finally, an incentive compatible contract with rent R must satisfy a modified version of (5.1) where rent R has been added to the right hand sides of (5.1a) and (5.1c) and since the system is linear in utilities, we conclude $U_g^R(e) = U_g^0(e) + R$ and $U_b^R(e) = U_b^0(e) + R$. *Q.E.D.*

Note a key difference between Lemma 5.1 and the situation in the case of two effort levels: Here, *ex-ante* incentive compatibility determines the form of any incentive-compatible contract up to the *ex-ante* rent level; in the discrete case, the *ex-ante* incentive constraints admit many solutions. The explanation is the familiar one that there is "less slack" when the agent has the option of making very small deviations. Note also that Lemma 5.1, which uses only the *ex-ante* incentive constraints, does not determine the contract offered to $\underline{e}$. In both the commitment and renegotiation cases, the cost-minimizing contract will offer $\underline{e}$ a riskless contract, so that $U_g^R(\underline{e}) = U_b^R(\underline{e}) = D(\underline{e}) + R$.

For future use, we denote by $e^*(R)$ the optimal effort for the principal to induce under commitment given the incentive constraint and that the agent obtain rent R:

$$e^*(R) = \arg\max_{e \in E} \left[\Pi^R(e) \right], \tag{5.2}$$

where

$$\Pi^R(e) \equiv p(e)g + (1-p(e))b - p(e)\Phi\left(U_g^R(e)\right) - (1-p(e))\Phi\left(U_b^R(e)\right). \tag{5.3}$$

We assume that, for all $R \geqslant 0$, $\Pi^R(e)$ is strictly quasi-concave (so that the constrained optimum $e^*(R)$ is unique) and that $e^*(R)$ strictly exceeds $\underline{e}$. Since effort is unbounded above, the principal strictly prefers $e^*(R)$ to any larger effort level. We assume that $\Pi^R(+\infty) \leqslant \Pi^R(\underline{e})$.

We will also use the following assumption.

ASSUMPTION A: *The marginal cost of the agent's rent for the principal*:

$$\left[p(e)\Phi'\left(U_g^R(e)\right) + (1-p(e))\Phi'\left(U_b^R(e)\right) \right],$$

is increasing with effort.

A sufficient condition for Assumption A to hold is that $\Phi''' \geqslant 0$,[7] which is satisfied when the agent has constant absolute or relative risk aversion. This assumption is used only to obtain the specific form of the optimal distribution (equation (5.5) below) and is not needed for any of the qualitative results.

Let $M(c, F)$ denote the expected wage bill to contract c under distribution $F(e)$.

DEFINITION 5.2: A contract menu c that is ex-ante incentive compatible with rent R for support E^* is *renegotiation-proof for distribution F with support E^** if it solves

$$\min M(\tilde{c}, F) \quad \text{such that} \tag{5.4}$$

$$p(e)\tilde{U}_g(e) + (1-p(e))\tilde{U}_b(e) \geqslant D(e) + R, \quad \forall e \in E^*, \tag{5.4a}$$

$$p(e)\tilde{U}_g(e) + (1-p(e))\tilde{U}_b(e) \geqslant p(e)\tilde{U}_g(\tilde{e}) + (1-p(e))\tilde{U}_b(\tilde{e}), \quad \forall e, \tilde{e} \in E^*. \tag{5.4b}$$

Condition (5.4a) is the interim IR constraint, and (5.4b) is the constraint that each type report truthfully. Since the incentive-compatibility constraints determine all of the terms of a rent-R contract except for the contract offered $\underline{e}$, the main force of the no-renegotiation constraint is to restrict the admissible set of distributions.

[7] To prove this, use the facts that $U_g^R > 0$ and that $p(e)\dot{U}_g^R + (1-p(e))\dot{U}_b^R = 0$ (both resulting from (5.1a) and (5.1b)).

LEMMA 5.2: *The set $P(R, F)$ of renegotiation-proof contracts for F with rent R is either empty or contains the single element defined by system* (5.1) *and, if $\underline{e} \in E^*$, the constraint $U_g(\underline{e}) = U_b(\underline{e}) = D(\underline{e}) + R$. Moreover, the renegotiation-proof contract satisfies*

$$p(e_1)U_g^R(e_1) + (1-p(e_1))U_b^R(e_1)$$
$$> p(e_1)U_g^R(e_2) + (1-p(e_1))U_b^R(e_2),$$

for all $e_1 \neq e_2 \in E^*$. That is, in the interim stage every type strictly prefers to report truthfully.

PROOF: The first part of the lemma follows immediately from Proposition 4.1 and the fact that the cost-minimizing compensation scheme to induce $\underline{e}$ is the full-insurance scheme $U_g(\underline{e}) = U_b(\underline{e}) = D(\underline{e}) + R$.

The final inequality follows from the fact that the agent's *ex-ante* payoff is strictly concave in his effort, so that

$$p(e_1)U_g^R(e_1) + (1-p(e_1))U_b^R(e_1)$$
$$= R + D(e_1)$$
$$= p(e_2)U_g^R(p_2) + (1-p(e_2))U_b^R(e_2) + D(e_1) - D(e_2)$$
$$> p(e_1)U_g^R(e_2) + (1-p(e_1))U_b^R(e_2). \qquad \text{Q.E.D.}$$

Given that the set $P(R, F)$ has at most one element, the key part of the characterization of optimal renegotiation-proof contracts is to determine the optimal distribution F for each rent R, which we do in Theorem 5.1. This step was trivial in the two-effort case: the principal preferred that $x = \text{Prob}(e = \bar{e})$ be as large as possible. In the continuum case the principal will prefer distributions that put as much weight on efforts near $e^*(R)$ as is consistent with no renegotiation.

As in the discrete case, the renegotiation-proof constraint influences the optimal contract only through the choice of distribution: for a given rent, the contract used to implement a given distribution is the same in the commitment and renegotiation cases. Here, though, this identity is trivial because there is a unique way to implement a given distribution in the commitment case.

THEOREM 5.1: (i) *If distribution F is renegotiation-proof for some choice of initial contract, then F has a continuous nonzero density on $[\underline{e}, \tilde{e}]$ for some $\tilde{e} \geq \underline{e}$ and no atom except possibly at $\underline{e}$.*

(ii) *The highest payoff for the principal in any equilibrium is $\Pi(\hat{e}(R))$, where $\hat{e}(R) > e^*(R)$. This payoff is attained by a distribution whose support is $[\underline{e}, \hat{e}(R)]$; the distribution has an atom at $\underline{e}$ if and only if $\dot{D}(\underline{e}) > 0$. Under Assumption A, the*

density is given by the following "generalized hazard-rate condition:"

$$\frac{p(e)(1-p(e))}{\dot{p}(e)}\left[\Phi'\left(U_g^R(e)\right)-\Phi'\left(U_b^R(e)\right)\right]f(e) \tag{5.5}$$
$$=\int_{\underline{e}}^{e}\left[p(\tilde{e})\Phi'\left(U_g^R(\tilde{e})\right)+(1-p(\tilde{e}))\Phi'\left(U_b^R(\tilde{e})\right)\right]dF(\tilde{e}).$$

(iii) *There exists an equilibrium that attains the upper bound defined in* (ii).

SKETCH OF PROOF: The proof of Theorem 5.1 is lengthy, and has been placed in the Appendix. For those who prefer not to work through the details, we first provide some intuition for the result and then provide a detailed overview of the sequence of arguments involved.

To begin, we explain why a renegotiation-proof distribution must have connected support so that in particular the two-effort distributions of Section 2 are not renegotiation-proof when there is a continuum of effort choices. Fix two effort levels $\underline{e}<e_2$, and consider the menu $\{U_g^0(e),U_b^0(e)\}_{e=\underline{e},e_2}$ defined by (5.1a), (5.1b). Comparing this to the menu $\{\hat{U}_g^0(e),\hat{U}_b^0(e)\}_{e=\underline{e},e_2}$ for the two-effort model with $\bar{e}=e_2$, we see that incentive constraint (5.1b) differs from AIC in Section 2, reflecting the fact that in the continuum model the binding incentive constraint is the local one. As a result, under contract $\{U_g^0(e),U_b^0(e)\}$, the agent after choosing effort e_2 strictly prefers announcing e_2 to announcing $\underline{e}$ at the interim stage, while under $\{\hat{U}_g^0(e),\hat{U}_b^0(e)\}$, he is *indifferent* between the two announcements. Because $\{\hat{U}_g^0(e),\hat{U}_g^0(e)\}$ makes the agent indifferent between the two announcements at the interim stage, a new compensation scheme $\{\tilde{U}_g^0(e_2),\tilde{U}_b^0(e_2)\}$ that reduces type e_2's risk at the interim stage will also be attractive to type $\underline{e}$. This is why the renegotiation-proofness of two-effort menus in Section 2 depended on the relative probabilities of the two types. However, when $\{U_g^0(e),U_b^0(e)\}$ makes the agent *strictly* prefer to announce his true level of effort, type e_2's risk can be reduced at the interim stage without tempting type $\underline{e}$ to misreport its type. Thus we expect that in order for a contract to be renegotiation-proof in the continuum-of-efforts case, the distribution $F(e)$ must place enough weight on types "just under e_2," who would incur only a small loss from announcing they are type e_2, so that a nonnegligible improvement in the contract offered e_2 would induce "enough" of these types to misreport. The proof extends this intuition by first showing that any renegotiation-proof distribution has no "gaps" and also no "atoms." It then solves for the minimum probability weight on low efforts that is consistent with the principal not being able to gain by offering an alternative contract at the interim stage, which results in equation (5.5).

We now provide a more detailed sketch of the proof.

From Lemma 5.2 we know that, given R, the principal's only leeway is the choice of the support E^* and its distribution F so as to maximize $\int_{\underline{e}}^{\infty}\pi^R(e)\,dF(e)$. Although the principal would wish to put all the weight of the distribution at $e^*(R)$, this is incompatible with renegotiation-proofness. For any level of effort e, there must be a sufficiently high probability that the agent has chosen effort

$\tilde{e} < e$ to discourage the principal from offering more insurance to type e at the interim stage. This condition is reflected in equation (5.5), which says that at the optimal distribution the gain from giving more insurance to type e is exactly offset by the loss from increasing the utility of all types $\tilde{e} < e$. Suppose that at the interim stage the principal gives more insurance to types in $[e, e + de]$ while keeping type $(e + de)$'s utility constant: To the second order, we have:

$$p(e)\delta U_g(e) + (1 - p(e))\delta U_b(e) = 0, \tag{5.6}$$

with $\delta U_b(e) > 0$. This raises efficiency and the principal's welfare is increased by

$$\begin{aligned}\left[-p(e)\Phi'\left(U_g^R(e)\right)\delta U_g(e) - (1-p(e))\Phi'\left(U_b^R(e)\right)\delta U_b(e)\right] f(e)\, de \\ = (1-p(e))\left(\Phi'\left(U_g^R(e)\right) - \Phi'\left(U_b^R(e)\right)\right)\delta U_b(e) f(e)\, de.\end{aligned} \tag{5.7}$$

The types above $(e + de)$ do not want to choose the new contract for types in $[e, e + de]$, because the former value insurance less than the latter do (they have a higher probability of a good outcome) and they prefer not to choose the latter's initial contract. In contrast, the interim utility of types $\tilde{e} < e$ must be increased. One incentive compatible way of doing so is to increase all the $U_g(\tilde{e})$ and $U_b(\tilde{e})$ by the same uniform amount δV. Differentiating (5.1a) and using (5.1b), one has

$$\dot{V}(e) = \dot{p}(e)\left(U_g^R(e) - U_b^R(e)\right), \tag{5.8}$$

and hence, to preserve incentive compatibility at e,

$$\begin{aligned}\delta V &= -\delta\left(\dot{V}(e)\right) de = -\dot{p}(e)\left(\delta U_g(e) - \delta U_b(e)\right) de \\ &= \frac{\dot{p}(e)}{p(e)}\delta U_b(e)\, de.\end{aligned} \tag{5.9}$$

The associated cost to the principal is:

$$\left[\int_{\underline{e}}^{e}\left[p(\tilde{e})\Phi'\left(U_g^R(\tilde{e})\right) + (1-p(\tilde{e}))\Phi'\left(U_b^R(\tilde{e})\right)\right](\delta V) f(\tilde{e})\, d\tilde{e}\right]. \tag{5.10}$$

At the optimum, the probability weight on the low levels of effort should be as small as possible given the constraint that the gain from increased insurance not exceed the cost of the increased interim utilities for the low-effort types. That is, the expressions in (5.7) and (5.10) should be equal, which, together with (5.9), yields (5.5). Equation (5.5) is a generalized hazard rate condition that reflects the usual tradeoff in adverse selection models between a local increase in efficiency at a given type and an increase in all worse types' rent.[8] Indeed, if $\Phi(\cdot)$ were linear, (5.5) would yield an equation in which the distribution would enter only through the hazard rate $f(e)/F(e)$.

Equation (5.5) yields a first-order differential equation in the density f. The solution is determined up to a multiplicative factor: if F is a solution, ξF is also a solution (with a different upper bound $\hat{e}$). Intuitively, the optimal choice of ξ

[8] This is the relevant tradeoff in the "standard" case where the incentive constraints are upward-binding.

should put as much weight around $e^*(R)$ as is feasible, so that the upper bound of the support E^* should be above $e^*(R)$. The homogeneity of the solutions to (5.5) implies that the profit at $\hat{e}$ is equal to the average profit: by decreasing ξ by $d\xi$ around $\xi = 1$, the principal's profit changes by $-d\xi \int_{\underline{e}}^{\hat{e}} \pi^R(e)\, dF(e) + d\xi\, \pi^R(\hat{e}) = 0$ (where the second term reflects the shift of weight to (slightly increase) the upper bound of the distribution).[9]

This completes our sketch of the proof of Theorem 5.1, which characterizes the optimal renegotiation-proof contract for a given rent R. We now briefly investigate the optimal choice of the rent.

THEOREM 5.2: *Under Assumption A, if the agent's utility is logarithmic or exhibits constant absolute risk aversion, the principal leaves no rent to the agent, i.e.,* $R = 0$.

SKETCH OF PROOF: The analysis is similar to that of the discrete case. The case of a logarithmic utility ($\Phi(U) = \Phi(0)e^U$) is trivial under Assumption A:[10] The existence of a rent R multiplies both the left-hand and right-hand sides of (5.5) by e^R and therefore has no effect on the equation yielding the renegotiation-proof distribution, but reduces the principal's objective function.

The case of constant absolute risk aversion is straightforward but tedious. The method of proof is to differentiate (5.5) and obtain $\dot{f}(e) = N(e, R)f(e)$ for some function N. Direct computation shows that, for this class of utility functions, $\partial N/\partial R > 0$, so that increasing R lowers the right-hand side of the differential equation and lowers f for all e. Thus an increase in R both shifts the distribution of efforts toward low efforts,[11] and for a given distribution of efforts reduces the principal's objective function, and thus is not desirable. *Q.E.D.*

5B. *Implementation with a Single Contract and Uniqueness of the Equilibrium Payoff*

We now generalize the results of Section 3 and show that the renegotiation-proof solution, in which the agent chooses at the interim stage from a menu of contracts specified *ex-ante*, can alternatively be implemented by a *single* contract offered to the agent *ex-ante* and *renegotiated* after the effort is chosen. Moreover, when this single contract is offered, the resulting distributions of effort and compensation are unique.

As in the two-effort case, the single contract should (i) give the agent a rent of R if he chooses the highest desired level of effort, which here is $\hat{e}(R)$, and (ii) should give the agent a strictly lower rent if he chooses a lower level of effort. With more than two efforts there is an additional constraint: (iii) the agent

[9] If there were an upper bound on the feasible effort level, we would have to consider whether $\hat{e}(R)$ as we have defined it were feasible.

[10] The result actually holds even if Assumption A does not. This can be seen from equation (A.21). If $\{F, U_g^R, U_b^R\}$ is renegotiation-proof (satisfies (A.21) for some $\lambda(\cdot) \geq 0$), then $\{F, U_g^0, U_g^0\}$ is also renegotiation-proof (it satisfies (A.21) for $\lambda(\cdot)e^{-R} \geq 0$).

[11] Because $\hat{e}$ strictly exceeds $e^*(R)$, it is not quite straightforward that the rightward shift in the distribution increases welfare. A complete proof is available upon request from the authors.

should not be able to ensure a rent greater than R by choosing an effort *greater* than $\hat{e}(R)$.

In the two-effort case, the optimal commitment contract for effort $\bar{e}$ made the agent indifferent between $\underline{e}$ and $\bar{e}$, so to satisfy condition (ii) the single contract had to be riskier than the commitment contract for effort $\bar{e}$. With a continuum of efforts, the commitment contract for $\hat{e}(R)$ makes effort $\hat{e}(R)$ *strictly* optimal, so we can take $c_1 = \{U_g^R(\hat{e}(R)), U_b^R(\hat{e}(R))\}$ and satisfy conditions (ii) and (iii).

As in the two-effort case, the reason that the single contract gives uniqueness where the renegotiation-proof menu does not can be seen by considering whether $\text{Prob}(e = \underline{e}) = 1$ is a continuation equilibrium. With the renegotiation-proof menu, this is an equilibrium, as it gives the agent rent R. With the single risky initial contract, if the principal is certain that $e = \underline{e}$, he will give the agent a riskless contract which is just as good for type $\underline{e}$ as the initial one. This riskless contract will give $\underline{e}$ *less* rent than R, so the agent would strictly prefer to choose $\hat{e}(R)$ at the initial stage and then to refuse the proposed renegotiation, obtaining rent R.

Note that unlike the two-effort case here the principal *cannot* offer a contract that is on the same indifference curve for $\hat{e}(R)$ as c_1 and riskier, for these contracts will give rent greater than R to some types $e > \hat{e}(R)$.

As in Theorem 5.1, let $\hat{e}(R)$ be the upper bound on the support of the effort distribution that is optimal for a given rent R.

THEOREM 5.3: *Suppose that the agent accepts the single contract* $c_1 = \{U_g^R(\hat{e}(R)), U_b^R(\hat{e}(R))\}$. *In any Nash continuation equilibrium, the distribution of effort levels and payoffs are exactly the optima derived in Section* 5A.

PROOF: See Appendix 3.

COROLLARY: *The principal can implement the optimal renegotiation-proof solution by offering the single contract corresponding to* $\hat{e}(R^*)$, *where* R^* *is the rent in the optimal renegotiation-proof menu of contracts.*

6. DISCUSSION AND CONCLUSION

6A. *No Initial Contract*

In Section 4 we compared the equilibrium of our renegotiation model with that of the model where players can commit not to renegotiate. It is also instructive to consider the model where no initial contract can be signed. For this purpose assume that if the agent works but does not sign a contract with the principal, the agent can keep the income of the project; the contract is then a form of contingent sale. Thus, the agent should be thought of as an independent entrepreneur, whose only reason to contract with the principal is to obtain insurance. We assume that the principal can observe whether or not the agent chose to "produce," i.e., to exert effort, but as before cannot observe the level of effort chosen.

In this version of the model the agent's reservation utility is his maximal expected payoff when forced to bear all the project's risk by himself. If this maximum is obtained by the agent's not working, then it is easy to show that the unique equilibrium without initial contracts is for the agent to not produce.

The more interesting case is where the agent would choose to produce even without an insurance contract. The case of no initial contract, where the agent will bear all of the risk unless an interim contract is signed, is formally equivalent to the principal offering the single initial contract $w_g = g, w_b = b$. This is unlikely to be the optimal initial contract derived in Section 5, so typically there will be efficiency gains from introducing initial contracts even though these contracts may later be renegotiated.

6B. *Extensions*

Our model is the simplest modification of the standard principal-agent problem that captures the notion of renegotiation. We here discuss some of the features of more complex models.

Note that once the agent has chosen the compensation scheme $c(\bar{e})$ at the interim stage, there is an incentive for the two parties to renegotiate again (which in turn would be anticipated by the agent). Our assumption is that there is no time to renegotiate between the agent's choice of compensation scheme and the date at which the outcome is realized. More generally, we could consider a multi-stage bargaining game before the date of realization of outcome. For instance, if the principal makes k offers, our analysis is unchanged because "all the bargaining" would take place in the kth round, i.e., just before the outcome is observed. Note also that assuming that there are lags between offers is reasonable. For instance, the compensation committee of the board of directors may meet only once a year to discuss the CEO's compensation.

One interesting extension of our model would allow the date of realization of the outcome to be stochastic, with bargaining occurring during the time when the outcomes might be realized. Suppose for instance that the contract and the effort are chosen at date 0. The outcome arrives (once) at date $t = 1, 2, \ldots$ according to a Poisson process (for instance, it might be an invention, a breakthrough, or else a proof of unreliability of the product). At the beginning of each period, the principal can make a renegotiation offer. Because of the uncertain arrival date, the problem is similar to a multi-period adverse selection problem with commitment and renegotiation.[12]

[12] An obvious guess (which we have not tried to verify) is that the solution has the following "Coasian form" (see Hart-Tirole (1988)). Instead of fully separating as in the deterministic outcome model, the two types separate progressively over time. Each period the agent chooses between two compensation schemes that take effect if the outcome is realized in the period. One is safe (offers full insurance at level $\underline{U}(t)$) and the other risky ($U_g(t) > U_b(t)$). Type $\bar{e}$ always chooses the risky contract. Type $\underline{e}$ randomizes between the safe and the risky contract; once he has chosen a safe contract, he reveals his type and can pick only the safe contracts forever after. Conditional on the agent's having chosen the risky contracts up to date t, the principal believes $x(t)$ that the agent has type $\bar{e}$ grows over time and reaches 1 in finite time (at which date the risky contract becomes a safe one). The intuition behind this kind of dynamics is that each period, there is a risk for type $\underline{e}$ of being caught without insurance if he chooses the risky contract that period.

This model presents interesting features but seems hard to analyze. A simpler extension of our model has the outcome being revealed in two periods, one before renegotiation and one afterwards (that is, there is no time for renegotiation before the first outcome, but there is between the two outcomes). This model shows that our result that renegotiation-proof contracts are the commitment solution (for the given distribution and rent) is of limited generality: If both periods' outcomes are equally informative, the optimal commitment contract will weight both periods equally, while the optimal renegotiation-proof contract seems likely to put more weight on the first period (pre-renegotiation) outcome.

6C. *Executive Compensation*

While our paper focuses on the technical implications of the renegotiation-proofness constraint, our results are at least suggestive of explanations for some of the observed details of compensation contracts. For instance, our model has two simple implications for executive compensation:

(a) An executive who has made important long-run decisions (project or product choices, investments), will be offered the discretion to choose from a menu of compensation schemes, some offering a fairly certain payment and some offering a riskier, performance-related payment.

(b) An executive's compensation may be insensitive to how well the firm performs after he retires, even if this performance conveys important information about the executive's actions. More precisely, our theory predicts that we will observe a distribution of contracts, some of which depend on post-retirement performance and others of which do not.

Our casual impression is that real-world executive compensation schemes are consistent with this theory. As a very rough description, there are three items in executive compensation: salary (fairly independent of performance); earnings-related items (bonus and performance plans); and stock-related items (stock appreciation rights and phantom stock plans).[13]

Concerning (a), we note that executives seem to have a fair amount of discretion in choosing the riskiness of their compensation. Many managerial contracts specify that part or all of bonus payments can be transformed into stock options (or sometimes into phantom shares), either at the executive's discretion or by the compensation committee (presumably) at the executive's request. This operation amounts to transforming a safe income (the earned bonus) into a risky one tied to future performance.

A related feature of compensation plans is that stock options and stock appreciation rights, which can be exercised at any time between the issue date

[13] See Smith-Watts (1982) for a good survey of executive compensation. Bonus plans yield short-term rewards tied to the firm's yearly performance. Rewards associated with performance plans (which are less frequent and less substantial than bonus plans) are contingent on three-to-five years earning targets. Stock appreciation rights are similar to stock options and are meant to reduce the transaction costs associated with exercising options and selling shares. Phantom stock plans credit the executive with shares and pay him the cash value of these shares at the end of a prespecified time period.

(or a year or eighteen months after the grant of the option) and the execution date, are much more popular than restricted or phantom stock plans, which put restrictions on sale: in 1980, only 14 of the largest 100 U.S. corporations had a restricted stock plan as opposed to 83 for option plans. Few had phantom stock plans, and in about half the cases, these plans were part of a bonus plan, and therefore were conditioned on the executive's voluntarily deferring his bonus.

Concerning (b), we note that long-term rewards such as stock options and performance plans are typically forfeited if the executive leaves the firm or is fired. However, some contracts do allow a retiring manager to qualify for bonuses after retirement. This is loosely consistent with our theory, which does not explain firings and thus cannot explain why firing and retirement would be treated differently.

Our model may also be useful for understanding the observed details of contracts in other sorts of agency relationships. For example, it yields another explanation (in addition to consumer moral hazard and bankruptcy) of why warranties are fairly limited in practice. We have in mind the case of a defense contractor for whom a given project is substantial relative to the firm's size and who therefore may be risk averse. The time involved in assessing the durability and reliability of the contractor's equipment leaves plenty of scope for mutually advantageous renegotiation of warranty provisions. Another example is the design of sharecropping contracts. Here there is a simple technological explanation for the assumption that much of the agent's effort is taken before the outcome is observed.

Of course, other models may yield similar predictions. For example, the fact that managers have discretion in choosing their compensation can be explained by the desire to match the compensation to exogenous differences in preferences. We point out that if these differences remain constant or evolve deterministically throughout the manager's tenure then the manager could make a once-and-for-all choice of compensation plan when he is hired, so that the manager's preferences must evolve stochastically in order to explain the use of menus late in his tenure.

Department of Economics, MIT, Cambridge, MA 02139, U.S.A.

Manuscript received May, 1988; final revision received November, 1989.

APPENDIX I: PROOF OF THEOREM 5.1

Fix a rent R. We first characterize incentive compatible allocations in Lemma A.1. Lemmas A.2 and A.3 then compute the optimal renegotiation-proof contracts for a given R.

For $e \in E^*$, let $V(e)$ be the interim utility $p(e)U_g^R(e) + (1-p(e))U_b^R(e)$. We extend the definition of the interim utility $V(e)$ to all efforts in E, even those which are not chosen in equilibrium. There are many incentive-compatible ways to prevent agents from announcing efforts e which are not in E^*; the simplest is to only allow announcements in E^*. Without loss of generality, we will restrict attention to such mechanisms. Thus, the interim utility of agents who chose efforts not in E^* is

$$V(e) \equiv \sup_{\tilde{e} \in E^*} \left\{ p(e)U_g^R(\tilde{e}) + (1-p(e))U_b^R(\tilde{e}) \right\}.$$

Corresponding to $V(e)$ we can define the limiting values of the outcome-contingent or *ex-post* utilities $U_g(e)$ and $U_b(e)$ by fixing a sequence $\{\tilde{e}^n\}$ that approximates the supremum of $V(e)$ and setting $U_g(e) \equiv \lim_{n\to\infty} U_g(\tilde{e}^n)$ and $U_b(e) \equiv \lim_{n\to\infty} U_b(\tilde{e}^n)$.

LEMMA A.1: *$V(e)$ is continuous, increasing, and almost everywhere differentiable in e, with $\dot{V}(e) = \dot{p}(e)[U_g(e) - U_b(e)]$ a.e. Also, the difference $[U_g(e) - U_b(e)]$ is increasing in e, and $U_g(e) \geq U_b(e)$.*

PROOF OF LEMMA A.1: $V(e)$ is continuous and increasing because $p(e)$ is continuous and increasing and $U_g^R(\tilde{e}) \geq U_b^R(\tilde{e})$ for all $\tilde{e}$ in E^*. A monotone function is differentiable almost everywhere. To obtain the derivative $\dot{V}(e)$ and the desired properties of $[U_g(e) - U_b(e)]$, we again follow standard lines: Fix any e_1, e_2 in E; we have

$$V(e_1) \geq p(e_1)U_g(e_2) + (1 - p(e_1))U_b(e_2), \tag{A.1a}$$

$$V(e_2) \geq p(e_2)U_g(e_1) + (1 - p(e_2))U_b(e_1). \tag{A.1b}$$

Using (A.1), we obtain

$$\begin{aligned}[p(e_2) - p(e_1)][U_g(e_2) - U_b(e_2)] &\geq V(e_2) - V(e_1) \\ &\geq [p(e_2) - p(e_1)][U_g(e_1) - U_b(e_1)],\end{aligned} \tag{A.2}$$

and taking $e_2 \to e_1$ in (A.2) yields the equation for $\dot{V}(e)$. Because $V(e)$ and $p(e)$ are increasing, (A.2) implies $U_g(e) - U_b(e)$ is nonnegative and $[U_g(e) - U_b(e)]$ is increasing in e. *Q.E.D.*

Now we characterize the set of distributions F that are renegotiation-proof for a fixed rent R. Lemma A.2 shows that a renegotiation-proof distribution has no gaps, no atoms except perhaps at $\underline{e}$, and that $\underline{e}$ is in the support of F.

LEMMA A.2: *A distribution F is renegotiation-proof only if* (1) *$F(e)$ is continuous at all $e > \underline{e}$,* (2) *$F(e) > 0$ for all $e > \underline{e}$, and* (3) *$F(e)$ is strictly increasing at e when $F(e) < 1$.*

PROOF OF LEMMA A.2: Let us first show $F(e)$ is strictly increasing. Suppose that there exists $e_1 < e_2$ such that $F(e_2 + \varepsilon) - F(e_1) > 0$ for all $\varepsilon > 0$ and $\lim_{e\to e_2^-} F(e) = F(e_1)$. Note that $e_2 \in$ support $F = E^*$, so that e_2 must be an optimal interim announcement for type e_2, and thus $U_g(e_2) = U_g^R(e_2)$ and $U_b(e_2) = U_b^R(e_2)$. At the interim stage, the principal can offer to change the *ex-post* utilities $\{U_g(e), U_b(e)\}$ into $\{U_g(e) + \delta U_g, U_b(e) + \delta U_b\}$ on $[e_2, e_2 + \varepsilon]$ and keep the same utilities for all other types, where δU_g and δU_b leave type $(e_2 + \varepsilon)$ with the same interim utility and satisfy $\delta U_b > 0$, so that all types in $[e_2, e_2 + \varepsilon]$ get more insurance in the new contract. We will show that the contract

$$\left\{\{U_g(e), U_b(e)\}_{e\notin[e_2, e_2+\varepsilon]}, \{U_g(e) + \delta U_g, U_b(e) + \delta U_b\}_{e\in[e_2, e_2+\varepsilon]}\right\} \tag{A.3}$$

(i) makes all types at least as well off as the initial contract, so that it is accepted, (ii) is incentive compatible, and (iii) yields a strictly higher welfare to the principal than the initial contract, so the initial contract is not renegotiation-proof.

(i) For almost all e in $[e_2, e_2 + \varepsilon]$, the equation for $\dot{V}(e)$ yields that the change in $\dot{V}(e)$ is negative. Since by construction, the change in $V(e_2 + \varepsilon)$ is zero, for e in $[e_2, e_2 + \varepsilon)$, the change in $V(e)$ is positive, which implies that the new contract is accepted.

(ii) For the new contract to be incentive compatible, it suffices that no type in E^* is tempted to take one of the compensation schemes not geared to him. Incentive compatibility in the old contract implies that for all e and $\tilde{e}$ in $[e_2, e_2 + \varepsilon]$:

$$p(e)U_g(e) + (1 - p(e))U_b(e) \geq p(e)U_g(\tilde{e}) + (1 - p(e))U_b(\tilde{e}), \tag{A.4}$$

and hence

$$\begin{aligned}p(e)\big(U_g(e) + \delta U_g\big) + (1 - p(e))\big(U_b(e) + \delta U_b\big) \\ \geq p(e)\big(U_g(\tilde{e}) + \delta U_g\big) + (1 - p(e))\big(U_b(\tilde{e}) + \delta U_b\big).\end{aligned} \tag{A.5}$$

So, the new allocation is incentive compatible for types in $[e_2, e_2 + \varepsilon]$ because they also prefer their compensation schemes to those types in $E^* - [e_2, e_2 + \varepsilon]$.

Let us show that the types e in $E_1 \equiv E^* \cap [\underline{e}, e_1]$ do not want to take one of the new compensation schemes. Lemma 5.2 and the fact that e_2 is *ex-ante* optimal for the agent imply that for $e \in E_1$

$$p(e)U_g(e) + (1-p(e))U_b(e) > p(e)U_g(e_2) + (1-p(e))U_b(e_2).$$

But, for all $\tilde{e}$ in $[e_2, e_2+\varepsilon]$,

$$p(e)U_g(e_2) + (1-p(e))U_b(e_2) \geq p(e)U_g(\tilde{e}) + (1-p(e))U_b(\tilde{e})$$

so that

$$\text{(A.6)} \qquad \forall e \in E_1, \quad p(e)U_g(e) + (1-p(e))U_b(e) > \max_{\tilde{e} \in [e_2, e_2+\varepsilon]} \left[p(e)U_g(\tilde{e}) + (1-p(e))U_b(\tilde{e}) \right].$$

Now the fact that E_1 is compact implies that for ε and δU_b sufficiently small, (A.6) also holds for the new allocation:

$$\text{(A.7)} \qquad \forall e \in E_1, \quad p(e)U_g(e) + (1-p(e))U_b(e) > \max_{\tilde{e} \in [e_2, e_2+\varepsilon]} \left[p(e)\left(U_g(\tilde{e}) + \delta U_g\right) + (1-p(e))\left(U_b(\tilde{e}) + \delta U_b\right) \right].$$

Next, we show that a type $e \in E^*, e > e_2 + \varepsilon$, does not want to announce a type in $[e_2, e_2+\varepsilon] \equiv E_2$. The intuition is that these types value insurance less than types in $[e_2, e_2+\varepsilon]$ and so are not attracted by the new schemes. Incentive compatibility of the initial contract implies that:

$$\text{(A.8)} \qquad p(e)U_g(e) + (1-p(e))U_b(e) \geq p(e)U_g(\tilde{e}) + (1-p(e))U_b(\tilde{e}).$$

Because $\delta U_b > 0$ and $p(e) > p(e_2+\varepsilon)$, $p(e)\delta U_g + (1-p(e))\delta U_b < 0$, and so (A.8) implies that

$$p(e)U_g(e) + (1-p(e))U_b(e) > p(e)\left[U_g(\tilde{e}) + \delta U_g\right] + (1-p(e))\left[U_b(\tilde{e}) + \delta U_b\right].$$

We have thus demonstrated that the new contract is incentive compatible.

(iii) Lastly we show that the new contract raises the principal's interim welfare, so that the initial contract was not renegotiation-proof. The change in the principal's welfare is given by

$$\text{(A.9)} \qquad \delta\left(\int_{\underline{e}}^{+\infty} \pi(e)\, dF(e)\right)$$

$$= -\int_{e_2}^{e_2+\varepsilon} \left[p(e)\Phi'\left(U_g^R(e)\right)\delta U_g + (1-p(e))\Phi'\left(U_b^R(e)\right)\delta U_b \right] dF(e)$$

$$= [\delta U_b]\left[\int_{e_2}^{e_2+\varepsilon} (1-p(e))\left(\frac{p(e)(1-p(e_2+\varepsilon))}{p(e_2+\varepsilon)(1-p(e))}\right) \times \left(\Phi'\left(U_g^R(e)\right) - \Phi'\left(U_b^R(e)\right)\right) dF(e)\right],$$

which, for ε small, is approximately

$$\delta\left(\int_{\underline{e}}^{+\infty} \pi(e)\, dF(e)\right) \simeq [\delta U_b]\left[(1-p(e_2))\left(\Phi'\left(U_g^R(e_2)\right) - \Phi'\left(U_b^R(e_2)\right)\right)\right]\left[F(e_2+\varepsilon) - F(e_2)\right].$$

This change in welfare is strictly positive because $U_g^R(e_2) > U_b^R(e_2)$.

This proof also shows that $F(e) > 0$ for all $e > \underline{e}$. (It suffices to take $e_1 = \underline{e}$ in the previous proof; note that it was not assumed that $e_1 \in E^*$ in this proof.)

The proof that there is no atom except possibly at $\underline{e}$ is very similar to the proof that there exists no gap. For, suppose that there exits an atom $\alpha = F(e_2) - \lim_{e \to e_2} F(e) > 0$ at $e_2 > \underline{e}$. Then, let the principal offer a new contract $\{U_g(e_2) + \delta U_g, U_b(e_2) + \delta U_b\}$ to type e_2 at the interim stage such that

$$\text{(A.10)} \qquad p(e_2)\delta U_g = -(1-p(e_2))\delta U_b < 0.$$

That is, the interim utility of type e_2 is kept constant, and he is given more insurance, so the new contract changes the allocation for only one type instead of for an interval as in the no-gap proof. As before, the allocation for types above e remains incentive compatible. Let

$$\Delta(e) \equiv \left[p(e)U_g^R(e) + (1-p(e))U_b^R(e)\right] - \left[p(e)\left(U_g(e_2)+\delta U_g\right) + (1-p(e))\left(U_b(e_2)+\delta U_b\right)\right],$$

denote a type e in E^*'s incentive to choose the old allocation over type e_2's new allocation. Using the equation for $\dot{V}(e)$, it is easily checked that the function Δ is strictly quasi-convex for $e > \underline{e}$. Now $\Delta(e_2) = 0$ by construction. And so by quasiconcavity $\Delta(e) > 0$ for $e > e_2$, and, for δU_b small, $\Delta(\underline{e}) > 0$ from Lemma 5.2. Hence, there exists a unique $\hat{e}(\delta U_b) < e_2$ such that $\Delta(\hat{e}(\delta U_b)) = 0$. Furthermore, $\hat{e}(\delta U_b)$ tends to e_2 when δU_b tends to 0. Types in $[\hat{e}(\delta U_b), e_2]$ abandon the old contract, and pool with e_2 at the new contract $\{U_g(e_2)+\delta U_g, U_b(e_2)+\delta U_b)\}$, whereas types $e < \hat{e}(\delta U_b)$ stick to their old allocation.

For $e \in [\hat{e}(\delta U_b), e_2]$ (recall that there is no gap between $\underline{e}$ and e_2), let

$$\delta U_g(e) \equiv U_g^R(e_2) + \delta U_g - U_g^R(e)$$

and

$$\delta U_b(e) \equiv U_b^R(e_2) + \delta U_b - U_b^R(e)$$

denote the changes in *ex-post* utilities. From the continuity of $U_g^R(\cdot)$ and $U_b^R(\cdot)$, $\delta U_g(e)$ tends to δU_g and $\delta U_b(e)$ tends to δU_b as e tends to e_2. The change in the principal's welfare is equal to

$$\text{(A.11)} \quad \delta\left(\int_{\underline{e}}^{+\infty} \pi(e)\,de\right) = \alpha[1-p(e_2)]\left(\Phi'\left(U_g^R(e_2)\right) - \Phi'\left(U_b^R(e_2)\right)\right)\delta U_b - \int_{\hat{e}(\delta U_b)}^{e_2}\left[p(e)\Phi'\left(U_g^R(e)\right)\delta U_g(e) + [1-p(e)]\Phi'\left(U_b^R(e)\right)\delta U_b(e)\right] dF(e).$$

The first term on the right-hand side of (A.11) is of order δU_b and strictly positive. The second term on the right-hand side is negligible relative to the first term, as the integrand is of order δU_b, and the weight of the distribution between the bounds of the integral, $F(e_2) - F(\hat{e}(\delta U_b))$, tends to 0 when δU_b tends to 0, because $\hat{e}(\delta U_b)$ tends to e_2. Thus, the principal's welfare could be increased at the interim stage if there were an atom at e_2. This completes the proof of Lemma A.3. *Q.E.D.*

This proves part (i) of Theorem 5.1. Now we turn to part (ii). We determine the *optimal* renegotiation-proof distribution for the principal for a fixed rent R.

LEMMA A.3: *There exists an optimal renegotiation-proof distribution with rent R for the principal, with density satisfying* (5.5) *on some interval* $(\underline{e}, \hat{e}(R)]$. *It satisfies* $F(\underline{e}) = 0$ *if* $\dot{D}(\underline{e}) = 0$, $F(\underline{e}) > 0$ *if* $\dot{D}(\underline{e}) > 0$ *and* $F(\hat{e}(R)) = 1$. *Furthermore,* $\hat{e}(R) > e^*(R)$.

PROOF OF LEMMA A.3: From Lemma A.2, we know that $F(\cdot)$ is strictly monotonic up to $\hat{e}$, and continuous except possibly at $\underline{e}$. Let $\hat{e} = \inf\{e | F(e) = 1\}$ ($\hat{e}$ can be finite or infinite), and let $dF(e) = f(e)\,de$, where $f(\underline{e})$ is the right-hand derivative of F at $\underline{e}$.

We will first find the distributions F such that the incentive-compatible contract $\{U_g^R(e), U_b^R(e)\}_{e\in[\underline{e},\hat{e}]}$ is renegotiation proof, and then investigate which of these distributions, if any, attains the maximum in the definition of the optimal renegotiation-proof contract. Recall that the interim utility of type $e \in E^*$ in an incentive-compatible contract is $D(e) + R$. Now suppose that the principal offers the new contract $\{U_g(e), U_b(e)\}_{e\in[\underline{e},\hat{e}]}$. This allocation yields the new *ex-post* utility

$$\text{(A.12)} \quad V(e) = p(e)U_g(e) + (1-p(e))U_b(e).$$

As in the proof of Lemma A.1, interim incentive compatibility requires that $V(\cdot)$ be continuous, increasing, and a.e. differentiable in $(\underline{e}, \hat{e}]$, with

$$\text{(A.13)} \quad \dot{V}(e) = \dot{p}(e)\left(U_g(e) - U_b(e)\right) \quad \text{a.e.}$$

Last, one can w.l.o.g. assume that the new contract is accepted by all types in $[\underline{e}, \hat{e}]$ (if not, replace $U_g(e)$ and $U_b(e)$ by $U_g^R(e)$ and $U_b^R(e)$ for those types who prefer the initial allocation), so

$$V(e) \geq D(e) + R. \tag{A.14}$$

Moreover, under our concavity assumptions, these necessary conditions can be shown to be sufficient: For any function $V(e)$ that satisfies (A.13) and the interim IR constraint (A.14) there exists a contract $\{U_g(e), U_b(e)\}$ that satisfies the interim IC and IR constraints. Thus the implementable allocation is renegotiation-proof if the associated contract $\{U_g^R(e), U_b^R(e)\}_{e \in [\underline{e}, \hat{e}]}$ maximizes

$$\int_{\underline{e}}^{\hat{e}} \left[\left(p(e)g + (1-p(e))b \right) - \left(p(e)\Phi\left(U_g(e)\right) + (1-p(e))\Phi\left(U_b(e)\right) \right) \right] f(e)\, de,$$

subject to the constraints (A.12), (A.13), and (A.14).

The Hamiltonian for this program is (omitting the parameter e):

$$H = \left(I(e) - p\Phi\left(U_g\right) - (1-p)\Phi(U_b) \right) f + \lambda(V - D - R)$$
$$+ \gamma\left(V - pU_g - (1-p)U_b\right) + \mu\left(\dot{p}\left(U_g - U_b\right) \right).$$

The state variable is V and the control variables U_g and U_b. We have:

$$\frac{\partial H}{\partial U_g} = 0 = -p\Phi'\left(U_g\right) f - \gamma p + \mu \dot{p}, \tag{A.15}$$

$$\frac{\partial H}{\partial U_b} = 0 = -(1-p)\Phi'(U_b) f - \gamma(1-p) - \mu \dot{p}, \tag{A.16}$$

$$\dot{\mu} = -\frac{\partial H}{\partial V} = \lambda + \gamma, \tag{A.17}$$

$$\lambda \geq 0. \tag{A.18}$$

Adding (A.15) and (A.16) yields

$$\gamma = -\left[p\Phi'\left(U_g\right) + (1-p)\Phi'(U_b) \right] f. \tag{A.19}$$

Therefore, from (A.15):

$$\mu = \frac{p(1-p)}{\dot{p}} \left[\Phi'\left(U_g^R\right) - \Phi'\left(U_b^R\right) \right] f. \tag{A.20}$$

Equations (A.17), (A.19), and (A.20) and the fact that the first-order condition must hold at the implementable allocation yield the fundamental equation:

$$\frac{p(1-p)}{\dot{p}} \left[\Phi'\left(U_g^R\right) - \Phi'\left(U_b^R\right) \right] f = \int_d^e \left[\left(p\Phi'\left(U_g^R\right) + (1-p)\Phi'\left(U_b^R\right) \right) dF - \lambda\, d\tilde{e} \right]. \tag{A.21}$$

Next, we claim that $d = \underline{e}$. For, compute (A.21) at $e = d$. One has $U_g^R(d) = U_b^R(d)$, which is possible only for $d = \underline{e}$.

Any solution f to (A.21) is continuous and differentiable on $(\underline{e}, \hat{e}]$. We now devote special attention to the class of solutions to (A.21) that correspond to a zero shadow price for the interim individual rationality constraint ($\lambda(e) = 0$ for all e). The intuition for doing so is the same as in the discrete case. If for some e, $\lambda(e) > 0$, the principal would be tempted to renegotiate the contract of type e, as the weight $f(e)\,de$ put on neighboring types could be shifted to higher efforts without jeopardizing renegotiation-proofness.

The class of solutions corresponding to (A.21) for $\lambda(\cdot) \equiv 0$ is given by

$$\frac{p(1-p)}{\dot{p}} \left[\Phi'\left(U_g^R\right) - \Phi'\left(U_b^R\right) \right] f = \int_{\underline{e}}^e \left[p\Phi'\left(U_g^R\right) + (1-p)\Phi'\left(U_b^R\right) \right] dF(\tilde{e}). \tag{A.22}$$

The class of solutions of (A.22) is indexed by the value $f(\underline{e})$. That is, if $f(\cdot)$ is a solution, $\tilde{f}(e) \equiv (\tilde{f}(\underline{e})/f(\underline{e}))f(e)$ is also a solution (with a different $\hat{e}$ in order to satisfy $\tilde{F}(\hat{e}) = 1$). Is there an atom at $\underline{e}$? We must consider two cases:

(a) If $\dot{D}(\underline{e}) = 0$, $\lim_{e \to \underline{e}^+}[\Phi'(U_g^R(e)) - \Phi'(U_b^R(e))] = 0$ from (5.1b) so that taking the limit of (A.22) as e tends to $\underline{e}$ is consistent with $F(\underline{e}) = 0$. Intuitively, if effort levels just above $\underline{e}$ are not very costly for the agent, he can be induced to take them without being subjected to much risk. In this case, the principal has little incentive to renegotiate with types near to $\underline{e}$.

We claim that a distribution with $F(\underline{e}) = \alpha > 0$ is dominated by a distribution $\tilde{F}$ with density $\tilde{f} = f/(1-\alpha)$ on the same interval $[\underline{e}, \hat{e}]$, but without an atom at $\underline{e}$. One clearly has $\tilde{F}(\hat{e}) = 1$. Furthermore, welfare increases by

$$\text{(A.23)} \qquad \left[\int_{\underline{e}}^{\hat{e}} \pi^R(e)\frac{f(e)}{1-\alpha}de\right] - \left[\int_{\underline{e}}^{\hat{e}} \pi^R(e)f(e)\,de + \alpha\pi^R(\underline{e})\right]$$
$$= \frac{\alpha}{1-\alpha}\left[\int_{\underline{e}}^{\hat{e}} \pi^R(e)f(e)\,de - (1-\alpha)\pi^R(\underline{e})\right].$$

Now, the principal *ex-ante* could have obtained profit $\pi^R(\underline{e})$ by offering only the contract $\{U_g^R(\underline{e}), U_b^R(\underline{e})\}$. (Because this contract offers full insurance, it is renegotiation proof.) Hence, either the no-incentive contract yielding $\pi^R(\underline{e})$ is optimal (which we will see is not the case), or for the optimal distribution,

$$\text{(A.24)} \qquad \pi^R(\underline{e}) \leqslant \alpha\pi^R(\underline{e}) + \int_{\underline{e}}^{\hat{e}} \pi^R(e)f(e)\,de.$$

But (A.24) implies that (A.23) is positive and thus that the optimal distribution, if it is not degenerate at $\underline{e}$, has no atom at $\underline{e}$.

(b) If $\dot{D}(\underline{e}) > 0$, the limit of the left-hand side of (A.22) as e tends to $\underline{e}^+$ is strictly positive from (5.1b); hence (A.22) implies that there must exist an atom α at $\underline{e}$ sufficiently large to prevent renegotiation of the risky schemes just above $\underline{e}$. To compute the minimum size $\underline{\alpha}$ of the atom at $\underline{e}$, let

$$U_g^R(\underline{e}^+) \equiv \lim_{e \to \underline{e}^+} U_g^R(e) \quad \text{and} \quad U_b^R(\underline{e}^+) \equiv \lim_{e \to \underline{e}^+} U_b^R(e).$$

Note that $\dot{D}(\underline{e}) > 0$ implies $U_g^R(\underline{e}^+) > U_b^R(\underline{e}^+)$. Furthermore, by continuity at $\underline{e}$:

$$\text{(A.25)} \qquad p(\underline{e})U_g^R(\underline{e}^+) + (1-p(\underline{e}))U_b^R(\underline{e}^+) = p(\underline{e})U_g^R(\underline{e}) + (1-p(\underline{e}))U_b^R(\underline{e}) = D(\underline{e}) + R.$$

We now apply the same argument as in the proof that the distribution has no gaps. Offering a little bit more insurance to types in $(\underline{e}, \underline{e} + de]$ with a change $(\delta U_g, \delta U_b)$ satisfying

$$\text{(A.26)} \qquad p(\underline{e} + de)\delta U_g = -(1 - p(\underline{e} + de))\delta U_b < 0,$$

raises the principal's profit by (approximately)

$$\text{(A.27)} \qquad (1-p(\underline{e}))\left[\Phi'\left(U_g^R(\underline{e}^+)\right) - \Phi'\left(U_b^R(\underline{e}^+)\right)\right]\delta U_b f(\underline{e})\,de,$$

from (5.7). But, from (5.8), it raises the rent of type $\underline{e}^+$ by

$$\text{(A.28)} \qquad \frac{\dot{p}(\underline{e})}{p(\underline{e})}\delta U_b\,de,$$

so that the cost of raising the rent of the (fully insured) atom at $\underline{e}$ is

$$\text{(A.29)} \qquad \left[\frac{\dot{p}(\underline{e})}{p(\underline{e})}\delta U_b\,de\right]\left[\Phi'(D(\underline{e}) + R)\right]\alpha.$$

The minimum size of the atom $\alpha = \underline{\alpha}$ is such that the expressions in (A.27) and (A.29) are equal:

$$\text{(A.30)} \qquad \left[\frac{p(\underline{e})(1-p(\underline{e}))}{\dot{p}(\underline{e})}\delta U_b\,de\right]\left[\Phi'\left(U_g^R(\underline{e}^+)\right) - \Phi'\left(U_b^R(\underline{e}^+)\right)\right]f(\underline{e}) = \underline{\alpha}\Phi'(D(\underline{e}) + R).$$

Note that (A.30) is the limit of (A.22) at $\underline{e}$. An interesting fact is that the minimum atom $\underline{\alpha}$ is proportional to $f(\underline{e})$. Note also that (A.30) contains the case $\dot{D}(\underline{e}) = 0$ as a special case (for which $\underline{\alpha} = 0$).

The same reasoning as for $\dot{D}(\underline{e}) = 0$ proves that, unless $\pi^R(\underline{e})$ is the optimal profit, putting more weight at $\underline{e}$ than $\underline{\alpha}$ cannot be optimal for the distribution indexed by $f(\underline{e})$.

Thus, whatever $\dot{D}(\underline{e})$, the optimal solution in the subclass of functions satisfying (A.22) must belong to the subclass of functions $\mathcal{F}$ indexed by $f(\underline{e})$ and satisfying (A.22) and (A.30). Fixing a $F(\cdot)$ (defined by $\underline{\alpha}$ and $f(\cdot)$) belonging to $\mathcal{F}$, $F_\xi(\cdot) = \xi F(\cdot)$ also belongs to $\mathcal{F}$ for any ξ in $(0, 1/\underline{\alpha})$, where $F(\cdot)$ corresponds to $\underline{\alpha}_\xi = \xi\underline{\alpha}$ and $f_\xi(\cdot) = \xi f(\cdot)$. There is a one-to-one correspondence between the half-line $[\underline{e}, +\infty)$ and $\mathcal{F}$: for any ζ there is an "upper bound" of $F_\xi, \hat{e}_\xi$ in $[\underline{e}, +\infty)$, defined by $\hat{e}_\xi = \inf\{e | F_\xi(e) = 1\}$; $\hat{e}_\xi$ is differentiable and strictly increasing in ξ, tends to $\underline{e}$ when ξ tends to $1/\underline{\alpha}$, and tends to $+\infty$ when ξ tends to 0. Note that the elements in $\mathcal{F}$ are ranked by the criterion of first-order stochastic dominance. Note also that the class $\mathcal{F}$ contains as an extreme element (for $\xi = 1/\underline{\alpha}$) the full-insurance, no-incentive contract yielding $\pi^R(\underline{e})$.

To find the best distribution in the subclass $\mathcal{F}$ for the principal, it suffices to solve

$$\max_{\xi \in (0, 1/\underline{\alpha})} \left\{ \xi\underline{\alpha}\pi^R(\underline{e}) + \int_{\underline{e}}^{e_\xi} \pi^R(e)\xi f(e)\, de \right\}. \tag{A.31}$$

An optimal ξ^* for (A.31) is strictly interior. Because $\pi^R(e) > \pi^R(\underline{e})$ just above $\underline{e}$, a slight decrease in ξ under $1/\underline{\alpha}$ increases expected profit beyond $\pi^R(\underline{e})$. Thus the full-insurance contract is not optimal. When ξ goes to 0, the maximand in (A.31) tends to $\pi^R(+\infty)$, which is lower than $\pi^R(\underline{e})$ by assumption.

Furthermore, $\hat{e}_{\xi^*} > e^*(R)$: Suppose first that $\hat{e}_{\xi^*} < e^*(R)$. Then, a small decrease in ξ raises $F_\xi(\cdot)$ in the sense of first-order stochastic dominance, which raises expected welfare as $\pi^R(\cdot)$ is strictly increasing on $[\underline{e}, e^*(R)]$. Second, suppose that $\hat{e}_{\xi^*} = e^*(R)$. The derivative of the principal's welfare with respect to ξ at $\xi = \xi^*$ is equal to:

$$\left[\underline{\alpha}\pi^R(\underline{e}) + \int_{\underline{e}}^{e^*(R)} \pi^R(e) f(e)\, de \right] + \xi^* \pi^R[e^*(R)] \frac{d\hat{e}}{d\xi} f[e^*(R)]. \tag{A.32}$$

But, $F_\xi(\hat{e}_\xi) = \xi\underline{\alpha} + \int_{\underline{e}}^{\hat{e}_\xi} \xi f(e)\, de = 1$ for all ξ implies that

$$\underline{\alpha} + \int_{\underline{e}}^{\hat{e}_\xi} f(e)\, de + \frac{d\hat{e}}{d\xi} \xi^* f(e^*(R)) = 0. \tag{A.33}$$

Substituting (A.33) into (A.32) yields the expression:

$$\underline{\alpha}\big(\pi^R(\underline{e}) - \pi^R(e^*(R))\big) + \int_{\underline{e}}^{e^*(R)} \big[\pi^R(e) - \pi^R(e^*(R))\big] f(e)\, de \tag{A.34}$$

which is negative by definition of $e^*(R)$. Hence, a small reduction in ξ, i.e. a slight increase in $\hat{e}$, strictly raises the principal's expected welfare.

The expression in (A.34) with $e^*(R)$ replaced by $\hat{e}_\xi$ must be equal to 0 for $\xi = \xi^*$. Multiplying it by ξ^* yields:

$$\pi^R(\hat{e}_{\xi^*}) = \int_{\underline{e}}^{\hat{e}_{\xi^*}} \pi^R(e)\, dF_{\xi^*}(e). \tag{A.35}$$

That is, the profit at the upper bound is equal to the expected profit.

Note that this reasoning, which does not make use of $\lambda \equiv 0$, holds for general renegotiation-proof distributions and not only those in $\mathcal{F}$. The optimal distribution is thus such that the profit at the upper bound is equal to the expected profit.

Last, we show that under Assumption A, the optimal renegotiation-proof distribution belongs to $\mathcal{F}$ (that is, the shadow value $\lambda(\cdot)$ can be taken equal to zero), and thus the solution satisfies (5.5). Suppose that the distribution G, with density g and atom β at $\underline{e}$, is optimal for the principal (G does not necessarily belong to $\mathcal{F}$). Let $\hat{e}$ denote the upper bound at G. We consider two cases.

Case 1: $\hat{e} \leqslant e^*(R)$. Let $F_\xi(\cdot)$ denote the distribution in $\mathcal{F}$ that has the same upper bound $\hat{e}$ as G. Note that $f_\xi(e) < g(e)$ for all e in $(\underline{e}, \hat{e}]$ is impossible, for if $f_\xi(\underline{e}) < g(\underline{e})$, then because the atom α_ξ for $F_\xi(\cdot)$ at $\underline{e}$ is minimal given $f_\xi(\underline{e})$, the atom β necessarily exceeds α_ξ (from A.21). Hence, F_ξ strictly first-order stochastically dominates G, which contradicts the fact that they have the same upper bound. Hence, there must exist some $e < \hat{e}$ such that $f_\xi(e) = g(e)$. Now, writing the

differentiable version of (A.21) for the two densities yields:

$$\dot{L}g + L\dot{g} = Mg - \lambda, \quad \text{and} \tag{A.36}$$

$$\dot{L}f_\zeta + L\dot{f}_\xi = Mf_\xi, \quad \text{where} \tag{A.37}$$

$$L(e) \equiv \frac{p(e)(1-p(e))}{\dot{p}(e)}\left[\Phi'\left(U_g^R(e)\right) - \Phi'\left(U_b^R(e)\right)\right], \tag{A.38}$$

$$M(e) \equiv p(e)\Phi'\left(U_g^R(e)\right) + (1-p(e))\Phi'\left(U_b^R(e)\right). \tag{A.39}$$

Because $\lambda \geqslant 0$, (A.36) and (A.37) imply that $\dot{f}_\xi(e) \geqslant g(e)$ when $f_\xi(e) = g(e)$. Hence, F_ξ dominates G in the sense of first-order stochastic dominance. Because $\pi^R(\cdot)$ is increasing on $[\underline{e}, \hat{e}]$, $\int_{\underline{e}}^{\hat{e}} \pi^R(e)\, dF_\xi(e) \geqslant \int_{\underline{e}}^{\hat{e}} \pi^R(e)\, dG(e)$ with strict inequality unless $G = F_\xi$.

Case 2: $\hat{e} > e^*(R)$. We must adjust the proof of Case 1 as $\pi^R(\cdot)$ is decreasing on $[e^*(R), \hat{e}]$. Define the distribution $F_\xi(\cdot)$ in $\mathscr{F}$ by $F_\xi(e^*(R)) = G(e^*(R)) < 1$. From the reasoning in Case 1, F_ξ first-order stochastically dominates G on $[\underline{e}, e^*(R)]$. So consider the distribution H defined by

$$H(e) = \begin{cases} F_\xi(e) & \text{for } e \leqslant e^*(R), \\ G(e) & \text{for } e \geqslant e^*(R). \end{cases} \tag{A.40}$$

The change in the principal's payoff is

$$\int_{\underline{e}}^{\hat{e}} \pi^R(e)(dH(e) - dG(e)) = \int_{\underline{e}}^{e^*(R)} \pi^R(e)\left(dF_\xi(e) - dG(e)\right) \geqslant 0, \tag{A.41}$$

with strict inequality unless $G = F_\xi$ up to $e^*(R)$. We will show that H is renegotiation-proof, and thus G cannot be optimal unless $G = F_\xi$ a.e. up to $e^*(R)$; and that $G = F_\xi$ a.e. beyond $e^*(R)$ as well.

H is, like F_ξ, renegotiation-proof up to $e^*(R)$. Would the principal want to offer more insurance to types $e > e^*(R)$? Because G is renegotiation-proof, and from (A.21) and $\lambda \geqslant 0$:

$$\frac{p(1-p)}{\dot{p}}\left[\Phi'\left(U_g^R\right) - \Phi'\left(U_b^R\right)\right] g \leqslant \int_{\underline{e}}^{e}\left[p\Phi'\left(U_g^R\right) + (1-p)\Phi'\left(U_b^R\right)\right] dG. \tag{A.42}$$

But Assumption A guarantees that the expression $[p\Phi'(U_g^R) + (1-p)\Phi'(U_b^R)]$ is increasing. Because H dominates G up to $e^*(R)$ and h coincides with g beyond $e^*(R)$, (A.42) implies for all $e \geqslant e^*(R)$:

$$\frac{p(1-p)}{\dot{p}}\left[\Phi'\left(U_g^R\right) - \Phi'\left(U_b^R\right)\right] h \leqslant \int_{\underline{e}}^{e}\left[p\Phi'\left(U_g^R\right) + (1-p)\Phi'\left(U_b^R\right)\right] dH. \tag{A.43}$$

Thus, the gain from insurance is lower than the cost of the increase in the agent's interim utility, so that H is indeed renegotiation proof.

Last, let us show that if G is optimal, $G = F_\xi$ a.e. for $e \geqslant e^*(R)$. Define the distribution $K(\cdot)$ with density $k(\cdot)$ by:

$$\begin{cases} K(e) = G(e) & \text{for } e \leqslant e^*(R), \\ \dot{L}k + L\dot{k} = Mk & \text{for } e > e^*(R). \end{cases} \tag{A.44}$$

Because $k(e^*(R)) = g(e^*(R))$, equations (A.44) imply that $k(e) \geqslant g(e)$ for all $e \geqslant e^*(R)$. Hence, beyond $e^*(R)$, G dominates K in the sense of first-order stochastic dominance. Because $\pi^R(\cdot)$ is decreasing in this range, K yields at least as much profit to the principal as G. Furthermore, K is renegotiation proof. We conclude that $G = F_\xi$ almost everywhere on $[\underline{e}, \hat{e}]$, so that the optimal solution belongs to $\mathscr{F}$ and satisfies (5.5). This concludes the proof of Theorem 5.1. *Q.E.D.*

Part (iii) follows along standard lines: Fix an initial contract c, and consider the minimization embedded in Definition 5.1, where the utility on the right-hand side of (5.4a) and (5.4b) are given by the initial contract. These constraints are linear in the utilities, and $M(\tilde{c}, F)$ is convex in utilities and linear in F. Thus the program has a unique solution, and the solution varies continuously in F (with the topology of weak convergence of measures, i.e., the weak-* topology). Also, the set of mixed best responses F for the agent is a nonempty valued, convex-valued, upper hemicontinuous function of the final contract c. Thus there exists a continuation equilibrium for every initial contract. *Q.E.D.*

APPENDIX 2: CONTINUA OF EFFORTS AND OUTCOMES

As the final check on the robustness of our results, we now consider the model with a continuum of efforts $e \in [\underline{e}, \infty]$ and a continuum of outcomes $y \in Y = [0, M]$; for notational simplicity we identify the outcome y with the principal's revenue (y). We maintain the same assumptions on the agent's utility function as before, and make the following assumptions on the distribution of outcomes: the cumulative distribution function $H(y; e)$ of y given effort e is differentiable, with continuous density $h(y; e)$. Moreover, the distribution satisfies the MLRP condition that $\partial/\partial y(h_e/h) \geq 0$, and the convexity condition $H_{ee} \geq 0$, where the subscript e denotes partial differentiation with respect to e. Grossman-Hart (1983) have shown that under these assumptions the agent's maximization problem is concave in e for any nondecreasing reward function $U(y)$, and that the cost-minimizing $w(y; e)$ that induces the agent to choose an effort $e > \underline{e}$ must be nondecreasing in y.

Under these assumptions we can show that the main conclusions of Section 5A carry through: A renegotiation-proof contract with rent R must involve the agent playing a mixed strategy with support on an interval $[\underline{e}, \hat{e}(R)]$; the optimal renegotiation-proof contract for a given distribution and rent is the same as the commitment one.

To prove this we can simply verify that each of the lemmas involved in the proof of Theorem 5.1 extends. We first note that Proposition 4.1 implies that the optimal renegotiation-proof contract for a given rent R and distribution F, if one exists, is simply the optimal incentive-compatible contract for the same rent and distribution.

Next we explain why once again any renegotiation-proof distribution must have a cumulative distribution function that is continuous and increasing, i.e., it has no atoms and no gaps. The proofs in the two-outcome case showed that if there were an atom at e_2 or a gap just below e_2, the principal could gain at the renegotiation stage by offering a new contract menu which was the same as the original one except on an interval $[e_2, e_2 + \varepsilon]$. The new menu provided more insurance to all of the types in the interval $[e_2, e_2 + \varepsilon]$, and held the utility of type $e_2 + \varepsilon$ constant. In the two-outcome case, the new menu was derived from the old one by slightly decreasing $U_g(e)$ and increasing $U_b(e)$ for all types in $[e_2, e_2 + \varepsilon]$. This new schedule clearly lowers the principal's wage bill; we also needed to argue that it would induce truthful revelation at the renegotiation stage, and in particular that types $e > e_2 + \varepsilon$ would not prefer to announce an effort $e \in [e_2, e_2 + \varepsilon]$. The intuition for this is that these high-effort types believe there is a greater probability of the good outcome, and so they are less attracted by an increase in U_b. Thus we could find a small change that improved the utility of types near e_2 but did not attract types with higher effort levels.

The analogous argument with a continuum of outcomes is that the principal can offer a new menu $\tilde{U}(y; e)$ such that $\tilde{U}(y; e_2 + \varepsilon)$ is a mean-preserving decrease in "risk" for type $e_2 + \varepsilon$. (We put "risk" in quotes because we are comparing distributions of utilities, and not distributions of wage payments.) Formally,

$$\int_0^M \tilde{U}(z; e_2 + \varepsilon) h(z; e_2 + \varepsilon)\, dz = \int_0^M U(z; e_2 + \varepsilon) h(z; e_2 + \varepsilon)\, dz$$

and

$$\int_0^y \tilde{U}(z; e_2 + \varepsilon) h(z; e_2 + \varepsilon)\, dz < \int_0^y U(z; e_2 + \varepsilon) h(z; e_2 + \varepsilon)\, dz \quad \text{for all} \quad 0 < y < M.$$

Because the schedules $U(y; \cdot)$ are increasing in y, and the agent is risk-averse, the new schedules lower the principal's expected wage bill. It remains to be shown that if the change is small enough, the new contract menu induces truthful revelation at the interim stage. This follows from our assumption of the MLRP condition, which implies that higher-effort types gain less from decreases in the risk of the utility distribution.

We conclude that as in the two-outcome case, any renegotiation-proof distribution must be continuous and have support on an interval $[\underline{e}, \hat{e}]$. Once again, the optimal choice of distribution for a given rent R will have $\hat{e}(R)$ greater than $e^*(R)$. We have not investigated the form the optimal distribution will take.

APPENDIX 3: PROOF OF THEOREM 5.3

Theorem 5.1 showed that a renegotiation-proof distribution has no gaps and no atoms. We now show this extends to any equilibrium distribution, and that the interim contract menu is exactly $\{U_g^R(e), U_b^R(e)\}_{e \in [\underline{e}, \hat{e}(R)]}$ as defined in Section 5A.

We will give only an outline of the proof, as the details follow the lines of the proof of Theorem 5.1. Let $F(e)$ be the equilibrium distribution of the agent's effort with upper bound $\bar{e}$, i.e., $\bar{e} = \inf\{e \mid F(e) = 1\}$. At the interim stage, the principal offers a menu of contracts $\{U_g(e), U_b(e)\}_{e \in E}$; without loss of generality, we can assume that each type e accepts the new contract. (If not, choose $\{U_g(e), U_b(e)\} = \{U_g^R(\hat{e}(R)), U_b^R(\hat{e}(R))\}$ for the types who would otherwise refuse the new offer.) The constraints are the interim IC and IR constraints:

$$\text{(A.45a)} \quad p(e)U_g(e) + (1-p(e))U_b(e) \geqslant p(e)U_g(e') + (1-p(e))U_b(e') \quad \text{for all } (e, e'),$$

$$\text{(A.45b)} \quad p(e)U_g(e) + (1-p(e))U_b(e) \geqslant p(e)U_g^R(\hat{e}(R)) + (1-p(e))U_b^R(\hat{e}(R)) \quad \text{for all } e.$$

As before, incentive compatibility implies that $[U_g(e) - U_b(e)]$ is a nondecreasing function of e.

Let us show that there is no gap in the distribution; i.e., $F(e)$ is strictly increasing for $e \leqslant \bar{e}$. To see this, suppose there is a gap between e_1 and $e_2 > e_1$. We claim that the interim IC constraint that type e_1 does not prefer to announce that it is e_2 must bind. If not,

$$p(e_1)U_g(e_2) + (1-p(e_1))U_b(e_2) = p(e_1)U_g(e_1) + (1-p(e_1))U_b(e_1) - \Delta,$$

where $\Delta > 0$ from (A.45a). Then the principal could increase his interim profit by offering the alternative contract

$$\tilde{c} = \left\{ \{U_g(e) - \Delta, U_b(e) - \Delta\}_{e \leqslant e_1}, \{U_g(e), U_b(e)\}_{e \geqslant e_2} \right\}$$

instead. Let us check that the new contract satisfies (A.45a) and (A.45b). Clearly, types above e_2 do not change their earlier choices. A type e below e_1 does not want to choose the contract of type $e' \leqslant e_1$ because all utilities under e_1 are reduced uniformly. So we must prove that a type $e \leqslant e_1$ does not want to choose the scheme of type $e' \geqslant e_2$. To see this, note that interim incentive compatibility of the original interim contract requires that the announcement-dependent utilities of type e (which are also denoted $\{U_g(e), U_b(e)\}$) satisfy $U_g(e) \geqslant U_b(e)$ and $U_g(e) - U_b(e)$ nondecreasing in e. Therefore, since under the new interim contract $\tilde{c}$, type e_1 is just indifferent between announcing e_1 and e_2, all of the types below e_1 would do strictly worse by reporting an $e \geqslant e_2$ than by reporting truthfully. Thus, the alternative contract is interim incentive-compatible.

Now since the alternative contract for e_2 is the same as the original one, which was individually rational, we know

$$\text{(A.46)} \quad p(e_2)U_g(e_2) + (1-p(e_2))U_b(e_2) \geqslant p(e_2)U_g(\hat{e}(R)) + (1-p(e_2))U_b(\hat{e}(R)).$$

Thus, since type e_1 is just indifferent under the alternative contract between announcing e_1 and e_2, and $p(e)$ is increasing, we know that the alternative contract satisfies (A.45b) for all $e \leqslant e_1$.

Thus, the alternative contract would be an improvement on the original one and so the interim IC constraint that type e_1 not announce he is e_2 must bind, i.e., $\Delta = 0$ in (A.46). But in this case the payoff from choosing effort e_1 at the *ex-ante* stage and announcing e_1 would strictly exceed that from choosing any effort $e \geqslant e_2$, so that we would have $F(e_1) = 1$, i.e., $e_1 = \bar{e}$, and the distribution would not have a gap.

Next, suppose that $F(e) = 0$ for all $e \leqslant e_0$, with $e_0 > \underline{e}$. Then, by the same reasoning as in the proof of Theorem 5.1, types around e_0 get (approximately) full insurance at the interim stage (no distortion at the bottom of the distribution). Hence, choosing $\underline{e}$ at the *ex-ante* stage and claiming effort e around e_0, yields strictly more utility than playing e and claiming e, which contradicts $F(e) = 0$ for all $e \leqslant e_0$. Thus $e_0 = \underline{e}$.

The above arguments show that the equilibrium distribution of efforts has support $[\underline{e}, \bar{e}]$ where $\bar{e} \leqslant +\infty$. The *ex-ante* equilibrium condition for the agent implies that for all e in $(\underline{e}, \bar{e}]$:

$$\text{(A.47a)} \quad p(e)U_g(e) + (1-p(e))U_b(e) - D(e) = R',$$

$$\text{(A.47b)} \quad p(e)\dot{U}_g(e) + (1-p(e))\dot{U}_b(e) = 0, \quad \text{a.e.}$$

where $R' \geqslant R$ in the IR constraint (A.47a) (because the agent can always choose effort $\hat{e}(R)$ and refuse to renegotiate) and (A.47b) is the first-order condition for the constraint that effort is chosen optimally given the agent's announcement. (We saw that incentive compatibility implies that $U_g(\cdot)$ and $U_b(\cdot)$ are monotonic, and thus a.e. differentiable.) Hence, almost everywhere on $[\underline{e}, \bar{e}]$, $\{U_g(e), U_b(e)\} = \{U_g^{R'}(e), U_b^{R'}(e)\}$.

Now we claim that the upper bound $\tilde{e} = \hat{e}(R)$. To show this, we first note that the interim IR constraint must bind for type $\tilde{e}$:

$$p(\tilde{e})U_g(\tilde{e}) + (1-p(\tilde{e}))U_b(\tilde{e}) = p(\tilde{e})U_g^R(\hat{e}(R)) + (1-p(\tilde{e}))U_b^R(\hat{e}(R)). \tag{A.48}$$

If not, the principal would reduce the wages of all types so as to reduce the interim utility by some $\Delta > 0$ and the contract would still satisfy the interim IR constraint by the argument used to prove the distribution has no gaps.

From Lemma 5.2, however, we know that choosing effort $\tilde{e}$ and announcing $\hat{e}(R)$ is strictly worse than choosing effort $\hat{e}(R)$ and reporting truthfully. Thus, if the $IR(\tilde{e})$ constraint binds as in (A.48), then the agent does strictly better by choosing $e = \hat{e}(R)$ and announcing $\tilde{e}(R)$ than choosing $e = \tilde{e}$. This implies $\tilde{e} = \hat{e}(R)$.

Now, it is clear that $R' = R$, for otherwise the principal could reduce all the utilities in the interim stage by $\Delta = R' - R$. Thus, the announcement-contingent payments and resulting support E^* of the agent's distribution of efforts is the same as that corresponding to the optimal renegotiation-proof contract of rent R.

We have now shown that the equilibrium distribution of efforts F has to support $[\underline{e}, \hat{e}(R)]$, and that the contracts offered will be exactly $\{U_g^R(e), U_g^R(e)\}$ as defined in Section 5A. It remains to show that the distribution of efforts is given by (5.5). Because renegotiation makes every type except $e = \hat{e}(R)$ better off than with the initial contract, the interim IR constraint does not bind. This means that when contemplating changes in contract that increase or decrease the risk faced by types in an interval $[e, e+de]$, the interim IC constraint is the only constraint the principal needs to consider.

The proof of Proposition 5.1 showed that even when the interim IR constraint binds, the right-hand side of (5.5), which is roughly a measure of the probability of types below e, must be at least as large as the probability of types near e. When the interim IR constraint does not bind, the principal can choose to *reduce* the insurance of types near e and *reduce* the interim utility of types below e. Thus in order for the principal's optimal interim proposal to be $\{U_g^R(e), U_b^R(e)\}$, equation (5.5) must hold with equality.

It should now be clear that the agent using distribution $F(e)$ given by Theorem 5.1 will be an equilibrium. *Q.E.D.*

REFERENCES

AGHION, P., M. DEWATRIPONT, AND P. REY (1989): "Renegotiation Design under Symmetric Information," MIT DP #522.

DEWATRIPONT, M. (1989): "Renegotiation and Information Revelation over Time: The Case of Optimal Labor Contracts," *Quarterly Journal of Economics*, 104, 589–620.

FUDENBERG, D., B. HOLMSTRÖM, AND P. MILGROM (1990): "Short-Term Contracts and Long-Term Agency Relationships," *Journal of Economic Theory*, 52, 194–206.

FUDENBERG, D., AND J. TIROLE (1988): "Moral Hazard and Renegotiation in Agency Contracts," MIT Discussion Paper #494.

GREEN, J., AND J. J. LAFFONT (1987): "Renegotiation and the Form of Efficient Contracts," mimeo, Harvard University.

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

HARSANYI, J. C. (1973): "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed Strategy Equilibrium Points," *International Journal of Game Theory*, 2, 1–23.

HART, O., AND J. MOORE (1988): "Incomplete Contracts and Renegotiation," *Econometrica*, 56, 755–785.

HART, O., AND J. TIROLE (1988): "Contract Renegotiation and Coasian Dynamics," *Review of Economic Studies*, 55, 509–540.

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

KREPS, D., AND R. WILSON (1982): "Sequential Equilibria," *Econometrica*, 50, 863–894.

LAFFONT, J. J., AND J. TIROLE (1988): "Adverse Selection and Renegotiation in Procurement," forthcoming, *Review of Economic Studies*.

MASKIN, E., AND J. MOORE (1988): "Implementation and Renegotiation," LSE mimeo.
MASKIN, E., AND J. TIROLE (1988): "The Principal-Agent Relationship with an Informed Principal, II: Common Values," MIT mimeo.
ROTHSCHILD, M., AND J. STIGLITZ (1976): "Equilibrium in Competitive Insurance Markets: An Essay in the Economics of Imperfect Information," *Quarterly Journal of Economics*, 90, 629–650.
SHAVELL, S. (1979): "Risk Sharing and Incentives in the Principal and Agent Relationship," *Bell Journal of Economics*, 10, 55–73.
SMITH, C., AND R. WATTS (1982): "Incentive and Tax Effects of Executive Compensation Plans," *Australian Journal of Management*, 7, 139–157.
STIGLITZ, J. (1977): "Monopoly, Non-linear Pricing and Imperfect Information: The Insurance Market," *Review of Economic Studies*, 44, 407–430.
WILSON, C. (1977): "A Model of Insurance Markets with Incomplete Information," *Journal of Economic Theory*, 16, 167–207.