working paper
department
of economics

MORAL HAZARD AND RENEGOTIATION
IN AGENCY CONTRACTS*

By Drew Fudenberg**
And
Jean Tirole**

No. 494 April 1988

massachusetts
institute of
technology
50 memorial drive
cambridge, mass. 02139

MORAL HAZARD AND RENEGOTIATION IN AGENCY CONTRACTS*

By Drew Fudenberg**

And

Jean Tirole**

No. 494 April 1988

* Research support from the National Science Foundation is gratefully acknowledged.

** Department of Economics, Massachusetts Institute of Technology

We would like to thank Bob Gibbons, Paul Healy, Oliver Hart, Jean-Jacques Laffont, and Mike Whinston for helpful comments.

MORAL HAZARD AND RENEGOTIATION IN AGENCY CONTRACTS*

By Drew Fudenberg**

and

Jean Tirole**

February 1988

Revised April 1988

* Research support from the National Science Foundation is gratefully acknowledged.

** Department of Economics, Massachusetts Institute of Technology

We would like to thank Bob Gibbons, Paul Healy, Oliver Hart, Jean-Jacques Laffont, and Mike Whinston for helpful comments.

ABSTRACT

We consider the problem of designing a contract between a risk-averse agent and a risk-neutral principal when the agent's action is subject to moral hazard and the principal is free to propose a new contract after the agent has chosen his effort level but before the corresponding outcome is revealed. In this setting any optimal contract is equivalent to one that is "renegotiation-proof." A renegotiation-proof contract that induces the agent to choose high effort levels by promising a higher payment following good outcomes must also induce the agent to choose lower effort levels with sufficiently high probability that the contract would not be renegotiated. We show that for a range of utility functions for the agent, including exponential and logarithmic forms, the cost-minimizing renegotiation-proof contract for a given distribution of efforts is the same as the cost-minimizing contract for that distribution under commitment. Thus, the force of the renegotiation-proof constraint is not to change the way that given distributions are implemented, but rather to change which distributions are feasible. However, if the agent has constant relative risk aversion lower than one, the principal may prefer to give the agent an ex-ante rent 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 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.

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. Holmstrom [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. Since the parties will rationally expect any contract that is not efficient at the interim stage to be renegotiated, any optimal contract is equivalent to one that is "renegotiation-proof." This paper studies the implications of the renegotiation-proofness constraint for the structure of optimal contracts.[1]

Fudenberg, Holmstrom and Milgrom [1987] show that the renegotiation-proofness constraint is not binding 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 Holmstrom [1979]) suggests that the agent's compensation should

---

[1]Recent studies of renegotiation-proofness in a Nash implementation context include Green-Laffont [1987], Hart-Moore [1986], Maskin-Moore [1987], and, in an adverse selection context, Dewatripont [1986], Hart-Tirole [1987] and Laffont-Tirole [1988].

then depend on the outcomes to be observed five years hence. However, since only the agent knows which action 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. 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. To simplify the analysis, we assume that the outcome has only two possible values, *g* and *b*. (We later show that our results still hold for general distributions of outcomes under the usual assumptions of monotone likelihood ratio property and convexity of distribution function.) After the effort is chosen, but before the results are 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. As a starting point we will 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". The bulk of the paper assumes that the agent's utility function is separable in income and effort; Section 3 discusses non-separable utility.

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 is 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 chose a lower level of effort (unless e^* is the lowest effort level.) This suggests that, if renegotiation is feasible, the equilibrium will be in mixed strategies: In order to make credible a contract that induces the agent to choose some $\hat{e} > 0$ 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 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.

At this point we should explain that any result that the principal can attain by some choice of ex-ante contract c_1 can be attained by a contract that is renegotiation-proof, i.e. that cannot be improved on at the renegotiation stage. To see this, imagine that c_1 is renegotiated to c_2, and consider offering c_2 as the *ex-ante* contract. We claim that c_2 must be renegotiation-proof: If there were a contract c_3 that upset c_2, it would have to offer at least as much expected utility to every type of the agent, and give a higher payoff to the principal. But then the principal would have chosen c_3 instead of c_2 when renegotiating the contract c_1, so c_2 is renegotiation proof.[2]

To understand the strucure of the optimal renegotiation-proof contract, it is helpful to follow Grossman and Hart [1983] and divide the principal's problem into parts. To solve the principal-agent model with commitment, Grossman and Hart use a three-step procedure. The first step is to characterize the set of incentive-compatible contracts that implement a given level of effort, or a given distribution over levels of effort. Next, find the element of this set that implements the desired distribution at the least cost to the principal. One expects that this least-cost contract will give the agent zero *ex-ante* rent, i.e. the agent's individual rationality constraint will bind. Finally, choose the distribution over effort that maximizes the difference between the principal's expected revenue and the cost of the agent's compensation.

To incorporate the renegotiation-proofness constraint, we subdivide Grossman and Hart's first step as follows. For a fixed distribution of efforts, we first characterize the set of incentive-compatible contracts ignoring renegotiation, and we then identify the subset which is renegotiation-proof. 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 is the same as the cost-minimizing contract under commitment. 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.

---

[2]Note that this argument relies heavily on the assumption that it is the principal who offers the contracts, so that the renegotiation is interim-efficient in the sense of Holmstrom-Myerson [1983].

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.

Sections 2 and 3 characterize the optimal renegotiation-proof contract menu when the agent can choose between two levels of effort. The agent randomizes between choosing a low effort and a fixed reward, and a high effort and a risky performance-related reward whose expected value exceeds that of the riskless option. The agent is given no *ex-ante* rent if his utility exhibits constant or increasing absolute risk aversion; he is given a rent if the utility exhibits constant relative risk aversion with coefficient strictly less than one <u>and</u> high performance is very valuable to the principal.

The results of Section 3 rely on the assumption 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 4 shows how to replace this assumption by expanding the model to include a small amount of private information about the agent's preferences. In the transformed model, the agent plays a pure strategy as a function of his preferences, and the overall distribution of effort levels corresponds to the mixed strategy we derived in Section 3. This result extends Harsanyi's [1973] observation that mixed strategies can be "purified" from games to mechanism design problems.

Section 5A extends the model to 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. Section 5B discusses the model with continua of efforts and outcomes.

Section 6 shows that the optimal renegotiation-proof contract, in which the agent chooses from a menu of incentive schemes, can 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.

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 which do not.

Our casual impression is that real-world executive compensation schemes are consistent with the 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).[3]

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 (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.

---

[3] 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.

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 potential 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.

2. The Model with Two Effort Levels and Two Outcomes

We begin by analyzing a simple model in which the agent has only two levels of effort and his performance can take two values. Section 5 considers the case of a continuum of effort levels and outcomes, and finds that the main results of this section extend quite naturally. 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 equal 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. There are two possible outcomes, *g* and *b*, with associated profits 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 $TR(e) = p(e)G + (1-p(e))B$ and the expected wage bill $E(w|e)$. The principal has the option of not employing the agent; we normalize this shut-down profit to be zero.

Before considering the structure of renegotiation-proof contracts it is helpful to examine the case where the parties can commit themselves not to renegotiate. While this analysis follows standard lines, it does introduce the new feature of implementing a non-degenerate distribution over effort levels. In the spirit of Grossman and Hart, we fix a desired distribution, with $x = \text{Prob}(e=\bar{e})$, and ask what contracts implement this distribution, and which contract is the cheapest way of doing so. To implement distribution x, the principal will offer the agent a menu of two contracts: contract $c(\underline{e})$ will induce the agent to take effort $\underline{e}$, and contract $c(\bar{e})$ will induce the agent to choose $\bar{e}$. The contracts specify money payments as a function of the realized outcome; we will find it convenient to describe the contracts by the utility the payments provide rather than by their money transfers. Thus contract $c(\bar{e})$ specifies two utility levels $\{U_g(\bar{e}), U_b(\bar{e})\}$, meaning that the principal will pay the agent $\Phi(U_g(\bar{e}))$ if outcome g occurs, and $\Phi(U_b(\bar{e}))$ following outcome b. Similarly, $c(\underline{e})$ specifies utilities $\{U_g(\underline{e}), U_b(\underline{e})\}$.

<u>Definition:</u> A contract menu $c = \{c(\underline{e}), c(\bar{e})\}$ is <u>incentive compatible with rent R for distribution x</u>, $0 < x < 1$, if it satisfies

(2.1) (a)
$$p(\bar{e})U_g(\bar{e}) + (1-p(\bar{e}))U_b(\bar{e}) - D(\bar{e}) - \\ p(\underline{e})U_g(\underline{e}) + (1-p(\underline{e}))U_b(\underline{e}) - D(\underline{e}) = R$$

(b)
$$p(\bar{e})U_g(\bar{e}) + (1-p(\bar{e}))U_b(\bar{e}) - D(\bar{e}) \geq \\ p(\underline{e})U_g(\bar{e}) + (1-p(\underline{e}))U_b(\bar{e}) - D(\underline{e})$$

(c)
$$p(\underline{e})U_g(\underline{e}) + (1-p(\underline{e}))U_b(\underline{e}) - D(\underline{e}) \geq \\ p(\bar{e})U_g(\underline{e}) + (1-p(\bar{e}))U_b(\underline{e}) - D(\bar{e}) \quad .$$

In system (2.1), condition (a) ensures that the agent is willing to accept both contracts, condition (b) is the incentive compatibility (IC) constraint for contract $c(\bar{e})$, and condition (c) is the IC constraint for contract $c(\underline{e})$. The term R is the utility rent the agent receives from accepting either contract. Note that these constraints are exactly those for the single contracts $c(\underline{e})$ and $c(\bar{e})$ each to be incentive compatible with rent R. These constraints guarantee that the agent will not prefer to "announce" an effort level different than the one he actually takes: Announcing effort e' when choosing effort $e \neq e'$ will give utility of no more than R. This property implies that optimal menus will be menus of the optimal contracts for single effort levels.

The principal's problem is to find the least-cost way of implementing distribution x. It will be useful to divide this problem in two: We first find the lowest-cost incentive-compatible contract that gives the agent a fixed rent of R, and then consider the optimal value of R.

<u>Definition</u>: The compensation cost of contract c with distribution x, is

(2.2)
$$M(c,x) = x[p(\bar{e})\Phi(U_g(\bar{e})) + (1-p(\bar{e}))\Phi(U_b(\bar{e}))] + (1-x)[p(\underline{e})\Phi(U_g(\underline{e})) + (1-p(\underline{e}))\Phi(U_b(\underline{e}))] \quad .$$

<u>Definition</u>: The cheapest incentive-compatible contract that implements x with rent R, $\hat{c}(R,x)$, solves

(2.3)
$$\text{Min } M(c(R,x),x)$$
$$\text{subject to } c(R,x) \text{ satisfies (2.1).}$$

In the minimization (2.3), constraint (2.1)(c), that contract $c(\underline{e})$ be incentive compatible, does not bind: The cheapest way to induce the agent to

take low effort while giving him rent R is to give a constant payment of $\Phi(D(\underline{e})+R)$, and it is easy to show that doing so is consistent with 2.1(c) whenever 2.1 (a) and (b) are satisfied.

Definition: The commitment solution for distribution x, $\hat{c}(x)$, minimizes the principal's expected payment (2.2) subject to (2.1) and the individual rationality constraint that $R \geq 0$.

Lemma 2.1: The commitment solution $\hat{c}$ for distribution x is independent of x and sets $R = 0$. The corresponding utility levels are:

$$
\begin{aligned}
(2.4) \qquad & U_g^o(\bar{e}) = [(1-p(\underline{e}))D(\bar{e}) - (1-p(\bar{e}))\, D(\underline{e})] \,/\, [p(\bar{e}) - p(\underline{e})] \\
& U_b^o(\bar{e}) = [p(\bar{e})D(\underline{e}) - p(\underline{e})\, D(\bar{e})]/[p(\bar{e}) - p(\underline{e})] \\
& U^o(\underline{e}) = D(\underline{e}).
\end{aligned}
$$

Further, for any rent $R \geq 0$, $\hat{c}(R,x) = \hat{c} + R$.

Proof: Standard.

Because the principal's payoff to distribution x is just the weighted average of her payoff when enforcing $e=\bar{e}$ and $e=\underline{e}$ as point distributions, the principal's preferences over contracts $\hat{c}(x)$ are monotonic in x.

3. Optimal Renegotiation-Proof Contracts with Two Effort Levels and Two Outcomes

Now we turn to the study of the optimal contract when the parties are free to renegotiate. We assume that the renegotiation takes the form of a take-it-or-leave-it offer by the principal. As we explained in the Introduction, we can without loss of generality restrict attention to contracts that are renegotiation-proof.

Imagine that the initial contract was $\{\{U_g(\underline{e}),U_b(\underline{e})\},\ \{U_g(\bar{e}),U_b(\bar{e})\}\}$ and that the principal believes the agent chose effort $\bar{e}$ with probability x. At the renegotiation stage the problem is one of insurance with adverse selection: it would be most efficient for the principal to offer the agent a deterministic wage schedule, but the principal does not know the value of this insurance to the agent because he does not know the agent's "type," i.e. his choice of effort.

<u>Definition</u>: A contract $c = \{U_g(\bar{e}),\ U_b(\bar{e}), U_g(\underline{e}),\ U_b(\underline{e})\}$ is <u>renegotiation proof for distribution x with rent R</u> if it solves (2.1) and also solves

(3.1)

$$\min_{\tilde{c}} M(\tilde{c},x)$$

subject to

(a) $$p(\bar{e})\tilde{U}_g(\bar{e}) + (1-p(\bar{e}))\tilde{U}_b(\bar{e}) \geq p(\bar{e})U_g(\bar{e}) + (1-p(\bar{e}))U_b(\bar{e}) - D(\bar{e})+R.$$

(b) $$p(\underline{e})\tilde{U}_g(\underline{e}) + (1-p(\underline{e}))\tilde{U}_b(\underline{e}) \geq p(\underline{e})U_g(\underline{e}) + (1-p(\underline{e}))U_b(\underline{e}) - D(\underline{e})+R.$$

(c) $$p(\underline{e})\tilde{U}_g(\underline{e}) + (1-p(\underline{e}))\tilde{U}_b(\underline{e}) \geq p(\underline{e})\tilde{U}_g(\bar{e}) + (1-p(\underline{e}))\tilde{U}_b(\bar{e})$$

(d) $$p(\bar{e})\tilde{U}_g(\bar{e}) + (1-p(\bar{e}))\tilde{U}_b(\bar{e}) \geq p(\bar{e})\tilde{U}_g(\underline{e}) + (1-p(\bar{e}))\tilde{U}_b(\underline{e}).$$

Conditions 3.1 (a) and (b) are the interim IR constraints for types $\bar{e}$ and $\underline{e}$, and (c) and (d) are the corresponding interim IC constraints.

Lemma 3.1: In an incentive compatible contract that is renegotiation-proof for some $x > 0$, $U_g(\underline{e}) = U_b(\underline{e}) = \underline{U}$. Furthermore, constraint 3.1(d) is not binding.

Proof: Consider the relaxed program of minimizing $M(\tilde{c},x)$ subject only to constraints 3.1 (a)(b) and (c). We will show that the solution satisfies 3.1(d) as well.

The solution to the relaxed program obviously satisfies $U_g(\underline{e}) = U_b(\underline{e}) = \underline{U}$: Adding risk to $\underline{e}$'s contract increases the cost of satisfying 3.1(c) but does not relax the other two constraints. Since the original contract induces the agent to choose $\bar{e}$ with probability $x > 0$, it must be that $U_g(\bar{e}) > U_b(\bar{e})$. This implies that 3.1(c) is satisfied with equality: Otherwise, more insurance could be given to $\bar{e}$ which would lower the cost of satisfying 3.1(a). From $U_g(\bar{e}) > U_b(\bar{e})$, $p(\bar{e}) > p(\underline{e})$, and the fact that 3.1(c) is satisfied with equality, we conclude that $p(\bar{e})U_g(\bar{e}) + (1-p(\bar{e}))U_b(\bar{e}) > p(\bar{e})U_g(\underline{e}) + (1-p(\bar{e}))U_b(\underline{e}) = \underline{U}$. Thus type $\bar{e}$'s IC constraint, 3.1(d), is satisfied. Q.E.D.

Lemma 3.2: When renegotiation is allowed, the probability x that the agent chooses the high effort must be strictly less than one.

Proof: If the probability of $\bar{e}$ is one, then any renegotiation-proof contract must give the agent a certain payment w; for individual rationality this payment must satisfy $U(w) \geq D(\bar{e})$. But then the agent would do better to choose $e=\underline{e}$, yielding $U(w)-D(\underline{e})>U(w)-D(\bar{e})\geq 0$. Q.E.D.

Definition: The set $P(R,x)$ is the set of contracts that are renegotiation-proof with respect to distribution x and give the agent a rent of R. If $P(R,x)$ is non-empty, $c^*(R,x)$ is the element that minimizes the principal's expected wage bill. Thus, $c^*(R,x)$ solves $\min_{c \in P(R,x)} M(c,x)$.

Lemma 3.2 shows that P(R,1) is empty for all R.

A necessary condition for a contract to be renegotiation-proof is that there not be a lower-cost contract that satisfies the constraints in (3.1) with equality. Despite the fact that these constraints are not the same as those in (2.1), the corresponding systems of linear equalities have the same solutions.

<u>Lemma 3.3:</u> If P(R,x) is nonempty, $c^*(R,x) = \hat{c}(R,x) = \hat{\underline{c}} + R$: A necessary condition for a contract with rent R to be renegotiation-proof is that it be the efficient way to give rent R in the commitment problem.

<u>Proof:</u> The contract that satisfies constraints 3.1 (a), (b), and (c) with equality solves $D(\underline{e}) + R = U(\underline{e})$, $D(\bar{e}) + R = p(\bar{e})U_g(\bar{e}) + (1-p(\bar{e}))U_b(\bar{e})$, and $p(\underline{e})U_g(\bar{e}) + (1-p(\underline{e}))U_b(\bar{e}) = U(\underline{e}) = D(\underline{e}) + R$. The first two equalities are exactly the definitions of R in 2.1(a), and the third is the equality version of the IC constraint 2.1(b). Thus the contract that solves the constraints in 3.1 with equality is exactly $\hat{c}(R,x)$. For the original contract to be renegotiation-proof, the alternative of $\hat{c}(R,x)$ must not be cheaper, and since $c^*(R,x)$ must satisfy 2.1, we conclude that if P(R,x) is not empty then $c^*(R,x) = \hat{c}(R,x) = \hat{c} + R$. Q.E.D.

Thus for fixed R and x the commitment and renegotiation solutions are identical. The solutions can differ in at most two ways: (i) the distribution x will be different, because in the renegotiation case the agent must randomize, and (ii) the rent R, which is zero in the commitment solution, may be positive. As we will see, the principal may choose to offer a positive rent in order to permit a higher level of x to be renegotiation-proof. While the analysis so far has held the distribution x fixed, the principal would prefer to have x be as large as possible if $\bar{e}$ is optimal in the commitment case. Lemma 3.4 below characterizes the highest value of x, $x^*(R)$, for which P(R,x) is nonempty. Since $c^*(R,x)=\hat{c}(R,x)$, as long as P(R,x) is not empty, the principal has no reason to choose an R > 0 if x^* is decreasing in R.

<u>Definition</u>: The critical level $x^*(R)$ is the highest value of R for which P(R,x) is non-empty.

<u>Lemma 3.4</u>: For all $R \geq 0$, $x^*(R)$ is the unique solution in [0,1] of

$$(3.2) \qquad \frac{x}{1-x} = \frac{\Phi'(U^o(\underline{e})+R)}{(\Phi'(U_g^o(\bar{e})+R)-\Phi'(U_b^o(\bar{e})+R))} \quad \frac{(p(\bar{e})-p(\underline{e}))}{p(\bar{e})(1-p(\bar{e}))} .$$

<u>Remark:</u> The utility levels U^o in (3.2) are those of the commitment solution for x. The right-hand side of (3.2) is inversely related to the gains from providing more insurance to type $\bar{e}$; note for example that if the agent were risk-neutral then the right-hand side of (3.2) would be infinite, and so we would have $x^* = 1$.

<u>Proof:</u> Fix a contract of the form $\hat{c}(R,x) = (U_g^o(\bar{e})+R, U_b^o(\bar{e})+R, U^o(\underline{e})+R)$ and consider the minimization problem at the renegotiation stage, (3.1). Figure 1 depicts constraints 3.1(a) and (c) as a function of the level of utility $\tilde{U}(\underline{e})$ that the new contract offers type $\underline{e}$. These two constraints intersect at the point

$$(3.3) \qquad \tilde{U}_b(\bar{e}) = [p(\bar{e})\tilde{U}(\underline{e}) - p(\underline{e})\hat{U}]/[p(\bar{e}) - p(\underline{e})]$$
$$\tilde{U}_g(\bar{e}) = [\hat{U} - (1-p(\bar{e}))\tilde{U}_b(\bar{e})]/p(\bar{e})$$

with $\tilde{U}_g > \tilde{U}_b$ and $\bar{U} = D(\bar{e}) + R$.

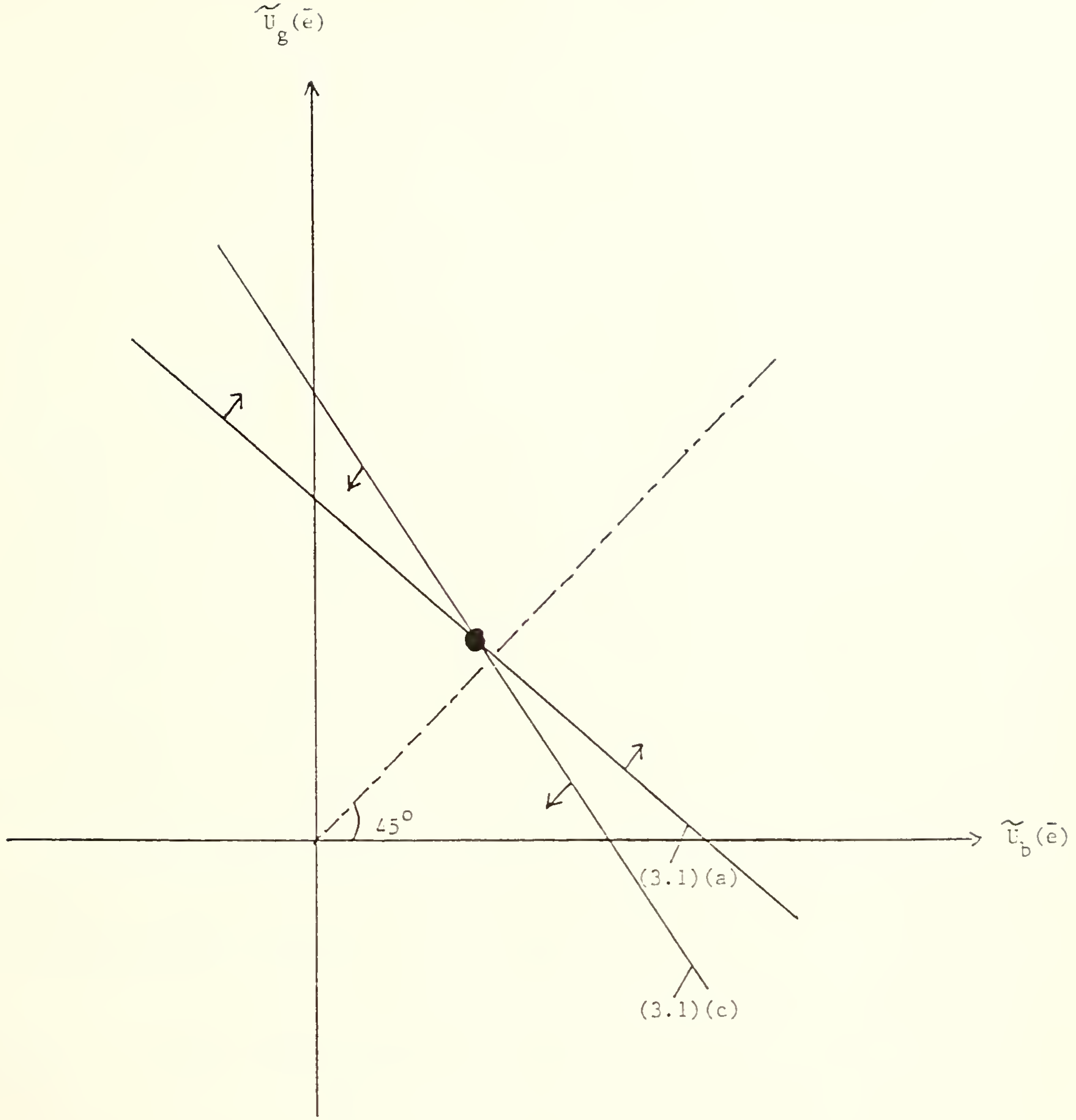

FIGURE 1

It is clear that 3.1(a), the IR constraint corresponding to $\bar{e}$, must bind, or the principal could reduce all of the utility levels and still satisfy the other constraints. This implies that 3.1(c), the IC constraint that type $\underline{e}$ not announce it is $\bar{e}$, should bind as well, for otherwise the principal could offer to lower his expected payments by offering type $\bar{e}$ a utility-preserving decrease in risk. However, constraint 3.1(b), that $\tilde{U}(\underline{e}) \geq U(\underline{e})$, need not bind: Increasing $\tilde{U}(\underline{e})$ relaxes the IC constraint 3.1(c) and allows the principal to reduce the risk faced by type $\bar{e}$. Whether doing this is worthwhile depends on the relation between the cost $\Phi'(U(\underline{e}))$ of raising type $\underline{e}$'s utility and the cost savings from reducing $\bar{e}$'s risk, which in turn depends on the amount of risk reduction, on the change in payments to type $\bar{e}$ for a given change in risk, and on the relative probability $x/(1-x)$ of the two types. If x is very large, the cost of increasing type $\underline{e}$'s rent is small, and the contract is less likely to be renegotiation-proof.

To obtain the equation for $x^*(R)$, we study how the principal's payoff at the renegotiation stage depends on the utility $v=\tilde{U}(\underline{e})$ that he offers type $\underline{e}$.

To simplify notation let $\tilde{U}_g(v) = \tilde{U}_g(\bar{e};v)$ and $\tilde{U}_b(v) = \tilde{U}_b(\bar{e};v)$ be the solution to 3.1(a) and (c) as given by (3.3), and let $M(v)$ be the corresponding expected cost to the principal. Then

(3.4) $$M(v) = x\,[p(\bar{e})\,\Phi\,(\tilde{U}_g(v)) + (1-p(\bar{e}))\,\Phi\,(\tilde{U}_b(v))] + (1-x)\,\Phi\,(v)$$

and

$$dM(v)/dv = -\,[xp(\bar{e})(1-p(\bar{e}))/(p(\bar{e})-p(\underline{e}))]\,[\Phi'(\tilde{U}_g(v))-\Phi'(\tilde{U}_b(v))] + (1-x)\,\Phi'(v)$$

where $\Phi'(U) = d\Phi/dU$, and where we have used (3.3) 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. Therefore at the solution to the minimization (3.1), $dM/dv \geq 0$, and either $dM/dv=0$ or $v=U(\underline{e})$. The condition $dM/dv \geq 0$ is exactly $x \leq x^*(R)$ as defined in

(3.2), where v is set equal to $R+D(\underline{e})$. Since the left-hand side of (3.2) is increasing in x for $x \in (0,1)$, if $dM/dv>0$ for some x and contract $\hat{c}(R,x)$, the same contract would be incentive-compatible and renegotiation proof for a larger x. Thus the highest x for which P(R,x) is nonempty is $x=x^*(R)$. Q.E.D.

The next step is to investigate how $x^*(R)$ varies with R. Inspecting the right-hand side of (3.2), we see that R influences $x^*(R)$ through its effect on the ratio

$$Q(R) = \Phi'(U^o(\underline{e})+R)/[\Phi'(U^o_g(\bar{e})+R)-\Phi'(U^o_b(\bar{e})+R)],$$

which compares the "demand for insurance" $[\Phi'(U^o_g(\bar{e})+R)-\Phi'(U^o_b(\bar{e})+R)]$ 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 3.5: (i) The sign of dx^*/dR is the negative of the sign of $d/dU\ (\frac{\Phi''(U)}{\Phi'(U)})$.

(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 $x \in (0,1)$, sign (dx^*/dR) = sign (dQ/dR). We compute

$$(3.5) \qquad \frac{dQ}{dR} \propto \Phi'(U^o_g(\bar{e})+R)\Phi'(U^o(\underline{e})+R)\left[\frac{\Phi''(U^o(\underline{e})+R)}{\Phi'(U^o(\underline{e})+R)} - \frac{\Phi''(U^o_g(\bar{e})+R)}{\Phi'(U^o_g(\bar{e})+R)}\right]$$

$$+ \Phi'(U^o(\underline{e})+R)\ \Phi'(U^o_b(\bar{e})+R)\ \left[\frac{\Phi''(U^o_b(\bar{e})+R)}{\Phi'(U^o_b(\bar{e})+R)} - \frac{\Phi''(U^o(\underline{e})+R)}{\Phi'(U^o(\underline{e})+R)}\right],$$

where $\propto$ means "proportional to". Since $U'' < 0$, $\Phi'' = \frac{-U''\cdot(\Phi')^2}{U'} > 0$. Since

$U_g^o(\bar{e}) > U(\underline{e}) > U_b^o(\bar{e})$, 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 $d(\Phi''/\Phi')/dU = \alpha[(\Phi''\Phi-(\Phi')^2)/\Phi^2] = [\alpha\Phi'/\Phi]\ [\Phi''/\Phi' - \Phi'/\Phi] = \alpha(\Phi'/\Phi)^2$ $(\alpha-1)$. Q.E.D.

The intuition for Lemma 3.5 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 $\Phi'(U(\underline{e})+R)$ of increasing $\underline{e}$'s utility. With increasing absolute risk aversion, increasing the rent <u>increases</u> 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 version 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.

<u>Definition:</u> The optimal renegotiation-proof contract is the choice of distribution x and renegotiation-proof contract $c^*(R,x)$ that maximizes the difference between the principal's expected revenue $xTR(\bar{e})+(1-x)TR(\underline{e})$ and the expected compensation cost.

<u>Assumption 3.1:</u> $TR(\bar{e})-[p(\bar{e})\Phi(U_g^o(\bar{e}))+(1-p(\bar{e}))\Phi(U_b^o(\bar{e}))] > \max\{0,\ TR(\underline{e})-\Phi(D(\underline{e}))\}$, so that in the commitment case the principal prefers to induce the agent to choose $\bar{e}$.

Lemma 3.6: Under Assumption 3.1, in the renegotiation case the principal will either shut down or will choose x and R so that $x = x^*(R)$.

Proof: This follows immediately from the observation that the commitment solution $\hat{c}(R,x)$ equals the renegotiation-proof solution $c^*(R,x)$. Q.E.D.

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 $TR(\bar{e})-TR(\underline{e})$ 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. In either case, the renegotiation-proofness constraint forces $x < 1$ and thus reduces the attractiveness of the contract to the principal, so it may be that the principal chooses to shut down when he would offer a contract in the commitment case.

Theorem 3.1: (i) If Φ''/Φ' is monotone decreasing or constant, the optimal renegotiation-proof contract has R=0. Thus either the principal shuts down or the contract is the same as in the commitment solution, except that the probability of high effort, x, is equal to $x^*(0)$ instead of 1.

(i) 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 $TR(\bar{e}) - TR(\underline{e}) - \Delta$. 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 Lemmas 3.5 and 3.6. The fact that the optimal contract is either shutdown or the same as the commitment solution follows from Lemma 3.3. In case (ii), note that the increase in expected revenue that the principal obtains by giving the agent a rent R>0 is $(x^*(R)-x^*(0))(TR(\bar{e})-TR(\underline{e}))$ which is increasing in $(TR(\bar{e})-TR(\underline{e}))$, so there is a Δ such that the principal gains by offering a nonzero rent if and only if $(TR(\bar{e})-TR(\underline{e}))>\Delta$. Assumption 3.1 implies that $\Delta>0$, and because the cost of offering a given rent R is finite, Δ is finite as well. Q.E.D.

Let us now briefly discuss the role of our assumption that the agent's utility is additively separable in income and effort. The argument that the commitment contract $\hat{c}(R,x)$ is the same as the renegotiation-proof contract $c^*(R,x)$ hinged on which of the IC and IR constraints were binding in each case. This argument is robust to small nonseparabilities in the agent's preferences, i.e. it will still hold for any utility function which is sufficiently close (in the C_o norm) to a separable function. Given that the same constraints bind, it is straightforward to show that $\hat{c}(R,x)= c^*(R,x)$: To see this, write the agent's utility as V(w,e). The optimal commitment contract, given that the IR constraint for effort $\bar{e}$ binds, and the IC constraint for type $\underline{e}$ is slack, satisfies

$$(3.6) \qquad R = E(V(\bar{w},\bar{e})|\bar{e}) = EV(\bar{w},\underline{e})|\underline{e}) = E(V(\underline{w},\underline{e})|\underline{e}),$$

where the first equality is the IR constraint for $\bar{e}$, the second is the *ex-ante* IC constraint for $\bar{e}$, and the last comes from the IR constraint for $\underline{e}$. These equalities determine three numbers $\bar{w}_g$, $\bar{w}_b$ and $\underline{w}$. The contract that satisfies the (equivalent of) renegotiation constraints (3.1) with equality, on the assumption that the IC constraint for $\bar{e}$ is slack, and the IR constraint for $\underline{e}$

binds, is

$$(3.7) \qquad E(V(\bar{w},\underline{e})|\underline{e}) = E(V(\underline{w},\underline{e})|\underline{e}) = E(V(\bar{w},\bar{e})|\bar{e}) = R,$$

where the first equality is the interim IC constraint for $\underline{e}$, and the last two are the IR constraints. Note that this is the same system of equalities as (3.6). Thus once again the optimal renegotiation-proof contract with zero rent is the same as the optimal commitment contract.[4] Of course, if there are substantial nonlinearities in the agent's preferences, then different constraints may be binding, and the two solutions need not be the same. 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 $\bar{e}$'s interim IC constraint.

4. Purification of Mixed Strategies

The fact that the optimal contract will involve the agent using a mixed strategy raises the following conceptual point: It is usual in agency models to assume that when the agent is indifferent between two pure actions he chooses the one that the principal most prefers. In our setting we have needed to make the stronger assumption that when the agent is indifferent he plays the mixed strategy that the principal prefers. The former assumption is typically justified by the observation that, at least if the agent is indifferent between two actions, a small change in the contract would make the agent prefer the desired action. Thus the tie-breaking rule simply means that

---

[4] Note that we restricted attention to deterministic contracts. Under non-separability, stochastic contracts may be optimal. But it is easy to see that such contracts cannot differ much from deterministic contracts for small non-separability.

the principal's set of feasible payoffs is closed. This argument on its own will not explain why the agent should randomize in the way that the principal would prefer. However, a mechanism-design version of Harsanyi's [1973] defense of mixed strategies will do the trick, as we now show. The idea is to introduce a little bit of uncertainty about the difference $D(\bar{e}) - D(\underline{e})$, 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 only for the two-action case, but we believe that versions of the result should obtain with more effort levels.

We start from the model of Section 2, and assume that the principal prefers the optimal renegotiation-proof contract to shutting down the firm. We consider a family of "elaborations" of the original model, indexed by n. In the each elaboration, the situation is the same as described in Section 2, except that the disutility of the high effort is $D(\bar{e}) + \epsilon$, where ϵ is private information to the agent before contracting. The private information ϵ is distributed according to the prior distribution $G^n(\epsilon)$ which is common knowledge. These elaborations are a small change to the original game in the sense that the family of distributions G^n converges to a point-mass at zero. That is, $\lim_{n\to\infty} (G^n(\epsilon) - G^n(-\epsilon)) = 1$ 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, $\lim_{n\to\infty} \bar{W}^n = W$, where W is the principal's maximized payoff as determined in Section 3, and $\bar{W}^n$ 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 $\bar{e}$.

Theorem 4.1: $\tilde{W}^n \to W$, and $\tilde{W}^n$ can be attained with a contract $\tilde{c}^n$ such G^n-almost all ϵ have a unique optimal action, and such that $\tilde{x}^n$ converges to the optimal level x obtained in Section 3.

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: We will first construct a sequence of renegotiation-proof contracts for the G^n, and show that they yield the principal a payoff that converges to W as $n \to \infty$. We then argue that these contracts are in fact optimal.

(i) Consider a contract menu that gives the agent two choices:

(a) a safe contract, yielding $U(\underline{e}) - D(\underline{e}) + R^*$ whatever the outcome (where R^* is the optimal rent in Section 3), and

(b) a risky contract $\{U_g^n(\bar{e}), U_b^n(\bar{e})\}$ defined by

$$(4.1) \qquad p(\bar{e})U_g^n(\bar{e}) + (1-p(\bar{e}))U_b^n(\bar{e}) - D(\bar{e}) - \epsilon_0^n = U(\underline{e}) - D(\underline{e}),$$

$$(4.2) \qquad \frac{G^n(\epsilon_0^n)}{1-G^n(\epsilon_0^n)} = \frac{p(\bar{e}) - p(\underline{e})}{p(\bar{e})(1-p(\bar{e}))} \; \frac{\Phi'(U(\underline{e})+R^*)}{[\Phi'(U_g^n(\bar{e})+R^*)-\Phi'(U_b^n(\bar{e})+R^*)]}.$$

$$(4.3) \qquad p(\underline{e})U_g^n(\bar{e}) + (1-p(\underline{e}))U_b^n(\bar{e}) = U(\underline{e}).$$

(4.1) says that type ϵ_0^n is indifferent between the two levels of effort. Types $\epsilon < \epsilon_0^n$ strictly prefer $\bar{e}$; from (4.2), the probability of these types is approximately the proportion x^* given by equation (3-2). The other types strictly prefer the low effort. Equation (4.3) is the interim IC constraint.

It is clear that $\lim_{n\to\infty} \epsilon_0^n = 0$, $\lim_{n\to\infty} G^n(\epsilon_0^n) = x^*$, and $\lim_{n\to\infty} (U_g^n(\bar{e}), U_b^n(\bar{e})) = (U_g^{R*}(\bar{e}), U_b^{R*}(\bar{e}))$. Thus the principal's payoff converges to W.

Now we must check that the allocation is renegotiation-proof. Since the agent's ex-ante private information concerns his disutility of effort, it is irrelevant at the interim stage. Thus the principal's problem at the interim stage is as in Section 3, and (4.2) ensures that the allocation is renegotiation-proof.

(ii) Conversely, let us show that the contract menu we specified is optimal for the principal subject to asymmetric information about ϵ and renegotiation-proofness. In general, the principal could use more complex menus of contracts, $(U_g^n(e,\epsilon), U_b^n(e, \epsilon))$ in which the agent announces his type ϵ as well as the effort e. However, since for each choice of effort e the agent's preferences over income are independent of ϵ, the principal can only offer two contracts $(U_g^n(\bar{e}), U_b^n(\bar{e}))$ and $(U_g^n(\underline{e}), U_b^n(\underline{e}))$ where $U_g^n(\underline{e}) = U_b^n(\underline{e}) = U(\underline{e})$ due to renegotiation-proofness. Ex-ante incentive compatibility implies that (4.1) holds for the type ϵ_0^n who is indifferent between the two effort levels. (4.2) and (4.3) then say that the contract is optimal for the principal subject to the renegotiation proofness and interim incentive compatibility constraints, respectively. Thus the optimal allocation under incomplete information about ϵ converges to the optimal contract under full information. Q.E.D.

Thus the mixed strategies called for by the optimal renegotiation-proof contract can be "purified" by introducing private information about the agent's disutility of effort. We have presented a proof only for one special model but we believe that the intuition will carry over to other contract problems in which the principal wishes to induce a mixed response from the agent, such as Laffont-Tirole [1988]. The intuition here is exactly that of the purification of mixed strategies in games: what looks like a mixed strategy may be the result of unobserved individual characteristics. The one new wrinkle involved in purifying mixed strategies in contracts as opposed to games is that the set of actions (i.e. contracts) that the principal may wish to consider is larger once unobserved characteristics are introduced. Under our assumption that the unobserved characteristics are an additional cost of effort which is added to the original utility function, these more complex contracts are not desirable.

5. Continuum Models

5A. Continuum of Efforts, Two Outcomes

We now generalize our earlier results to a continuum of effort levels. Section 5B discusses the case of continua of both efforts and outcomes. 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

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

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, ($\dot{D}(e) \geq 0$, $\ddot{D}(e) > 0$) 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 $V(e) = p(e)U_g(e)+(1-p(e))U_b(e)$ denote the agent's interim utility when he has chosen contract c(e) and effort level e.

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

As in the discrete case, the menu $\{U^R(e)\}$ is incentive compatible with rent R for a distribution on E^* iff for each $e' \in E^*$ the single contract $U^R(e')$ is incentive compatible with rent R. The following lemma shows that the incentive constraints and the rent R completely determine the form of these single contracts except at $\underline{e}$.

<u>Lemma 5.1:</u> A menu $\{U_g^R(e), U_b^R(e)\}$ is 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 $e \in E^*-\{\underline{e}\}$, where $\{U_g^0(e), U_b^0(e)\}$ is the unique solution on $E^*-\{\underline{e}\}$ of

(5.1)

$$\text{(a)}\quad p(e)U_g^0(e) + (1-p(e))U_b^0(e) - D(e) = 0 \qquad \forall e \in E^*-\{\underline{e}\}$$

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

$$\text{(c)}\quad p(e)U_g^0(e) + (1-p(e))U_b^0(e) - D(e) \le 0 \qquad \forall e \in E,$$

<u>Proof:</u> First, we claim any incentive compatible contract for E^* with rent 0 must satisfy system (5.1). Equation 5.1(a) simply says that the agent obtains

zero rent by choosing an effort in E^* and its corresponding contract; 5.1(c) says no choice of effort yields higher rent. 5.1(b) 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. Equations 5.1(a) and (b) are a non-singular linear system of two equations in two unknowns, and thus have a unique solution $(U_g^0(e), U_b^0(e))$. 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 (5.1) (a) and (c), 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, incentive compatibility determines the form of any incentive compatible contract up to the *ex-ante* rent level; in the discrete case, the 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 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 under full commitment under the constraint that the agent be given rent R:

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

where

$$(5.3) \qquad \Pi^R(e) \equiv p(e)G + (1-p(e))B - p(e)\Phi(U_g^R(e)) - (1-p(e))\Phi(U_b^R(e)).$$

We assume that, for all $R \geq 0$, $\Pi^R(e)$ is strictly quasi-concave (so that the constrained optimum $e^*(R)$ is unique) and that $e^*(R)$ strictly exceeds $\underline{e}$. We assume that $\Pi^R(+\infty) \leq \Pi^R(\underline{e})$.

We will also use:

<u>Assumption A:</u> The marginal cost of the agent's rent for the principal: $[p(e)\Phi'(U_g^R(e)) + (1-p(e))\Phi'(U_b^R(e))]$ is increasing with effort.

A sufficient condition for assumption A to hold is that $\Phi''' \geq 0$,[5] 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).

<u>Definition:</u> A contract menu c that is incentive compatible with rent R for support E^* is <u>renegotiation-proof for distribution F with support E^*</u> if it solves

(5.4) $\min M(c,F)$ such that

(a) $p(e)\hat{U}_g(e) + (1-p(e))\hat{U}_b(e) \geq D(e)+R \quad \forall e \in E^*$

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

Condition 5.4 (a) is the interim IR constraint, and 5.4 (b) is the constraint that each type report truthfully. Since the incentive-

---

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

compatibility constraints determine all of the terms of a rent-R contract except for the contract offered $\underline{e}$, the main force of the renegotiation-proof constraint is to restrict the admissible set of distributions.

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 the constraint $U_g(\underline{e}) = U_b(\underline{e}) = D(\underline{e})+R$.

Proof: Given the contracts for types $e > \underline{e}$ determined in Lemma 5.1, it is clear that the riskless contract $U_g(\underline{e}) = U_b(\underline{e}) = D(\underline{e})+R$ satisfies the interim incentive constraints and is cost-minimizing over the contracts that give the type $\underline{e}$ a rent of R. [Here and in the following, the contracts we discuss are to be understood as equivalence classes of contracts that agree almost everywhere with respect to distribution F.] Q.E.D.

Lemma 5.2 shows that, as in the two-effort case, the same contract is used to implement a given distribution and rent in both the commitment and renegotiation cases.

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 a given rent R. 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 choose distributions that put enough weight on efforts below and above $e^*(R)$ that it is too costly to offer more insurance at the interim stage.

Theorem 5.1, our main result for the continuum case, characterizes the principal's optimal choice of renegotiation-proof distribution for a fixed rent R. Later we discuss the optimal choice of the rent R, which, as in the two-effort case, depends on the rate at which the agent's risk aversion decreases with his income.

As in the discrete case, the renegotiation-proof constraint influences the optimal contract only through the choice of distribution: 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 contract in the commitment case.

Theorem 5.1: In the optimal contract, the agent receives rent $R \geq 0$ and plays a mixed strategy on $[\underline{e},\hat{e}(R)]$ where $\hat{e}(R) > e^{*}(R)$. The mixed strategy has a continuous density on $(\underline{e},\hat{e}(R)]$, and admits an atom at $\underline{e}$ if and only if $\dot{D}(\underline{e}) > 0$. The principal's expected profit is equal to $\Pi^{R}(\hat{e}(R))$. Last, under Assumption A, the density is given by the following "generalized hazard-rate condition":

$$(5.5) \qquad \frac{p(e)(1-p(e))}{\dot{p}(e)} [\Phi'(U_g^R(e)) - \Phi'(U_b^R(e))]f(e)$$

$$= \int_{\underline{e}}^{e} [p(\tilde{e})\Phi'(U_g^R(\tilde{e}))+(1-p(\tilde{e}))\Phi'(U_b^R(\tilde{e}))]dF(\tilde{e}).$$

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 now provide a detailed overview of the sequence of arguments involved.

A key step is to show that any renegotiation-proof distribution has connected support, i.e., there are no "gaps". Suppose to the contrary that the agent's strategy puts zero probability on a non-degenerate interval (e_1,e_2). Because $D(e_2) > D(e_1)$, the agent must face income risk after choosing e_2: If $U_g(e_2) = U_b(e_2)$, choosing effort e_1 and claiming effort e_2 would dominate choosing e_2 and claiming effort e_2. Thus, the principal could lower his wage bill by raising $U_b(e_2)$ and lowering $U_g(e_2)$. He may refrain from doing so, however, if incentive compatibility would require increasing the interim utility of other types as well. As in the discrete analysis the binding incentive compatibility constraints are the upward ones: at the interim stage the low- effort types want to claim that their effort was high. Now, if the contract was altered to give slightly less risk to types around e_2, the types $e \le e_1$ still would not prefer to claim they have type e_2, because they strictly preferred their scheme $(U_g(e), U_b(e))$ to the scheme $(U_g(e_2), U_b(e_2))$ from (5.1). Types $e > e_2 + \epsilon$ will also not be tempted to claim they are type e_2, because these types place less value on increases in U_b than type e_2 does. A similar proof shows that the lower bound of E^* must be $\underline{e}$, and that there cannot exist an atom in the distribution except possibly at $\underline{e}$.

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 on the point $e^*(R)$, this distribution is not renegotiation-proof. 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 princpal 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 those types' utility constant:

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

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

(5.7) $$[-p(e)\Phi'(U_g^R(e))\delta U_g(e) - (1-p(e))\Phi'(U_b^R(e))\delta U_b(e)]f(e)de$$

$$= (1-p(e))(\Phi'(U_g^R(e)) - \Phi'(U_b^R(e)))\delta U_b(e)f(e)de.$$

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 preferred 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 5.1 (a) and using 5.1 (b), one has

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

and hence, to preserve incentive compatibility at e,

(5.9) $$\delta V = -\delta(\dot{V}(e))de = -\dot{p}(e)(\delta U_g(e) - \delta U_b(e))de = \frac{\dot{p}(e)}{p(e)}\delta U_b(e)de.$$

The associated cost to the principal is:

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

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.[6] Indeed, if $\Phi(\cdot)$ were linear, (5.5) would yield a condition in 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 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).

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.

<u>Theorem 5.2:</u> Under assumption A, if the agent's utility is logarithmic or exhibits constant absolute risk aversion, the principal leaves no rent to the agent, $R = 0$.

<u>Sketch of Proof:</u> 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

---

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

assumption A:[7] 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 it 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,[8] and for a given distribution of efforts reduces the principal's objective function, and thus is not desirable. Q.E.D.

5B. 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 TR(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 F(y;e) of y given effort e is differentiable, with continuous density f(y;e). Moreover, the distribution satisfies the MLRP conditon that $\partial/\partial y\ (f_e/f) \geq 0$, and the

---

[7] The result actually holds even if assumption A does not. This can be seen from equation (A.25). If $\{F,U_g^R,U_b^R\}$ is renegotiation proof (satisfies (A.25) for some $\lambda(\cdot) \geq 0$), then $\{F,U_g^0,U_g^0\}$ is also renegotiation-proof (it satisfies (A.25) for $\lambda(\cdot)e^{-R} \geq 0$).

[8] 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.

convexity condition $F_{ee} \geq 0$, 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 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 monotonically increasing 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 has $\hat{e} > e^*(R)$, and 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 extend. We first note that once again, the optimal menu to induce a distribution over effort levels is a menu of the contracts that are optimal for each effort level considered separately. As in the case of two outcomes, the incentive contraints and rent level completely determine the optimal incentive-compatible contract for a given effort; the cost-minimization step is trivial. For this reason, it is again clear that the commitment and renegotiation-proof contracts for a given rent and distribution coincide. It remains to show that a distribution F is renegotiation-proof only if it has support on an interval $[\underline{e}, \hat{e}]$, and that the optimal choice of $\hat{e}$ for rent level R strictly exceeds the commitment level $e^*(R)$.

To do this we first note that a contract menu $\{U^R(y;e)\}$ is incentive compatible with rent R for support E^* iff for each e′ in E^*, $U^R(y;e')$ is the solution to the commitment problem of implementing e′ with rent R. 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 was the same as the original one except on an interval $[e_2, e_2+\epsilon]$. The new menu provided more insurance to all of the types in the interval $[e_2, e_2+\epsilon]$, and held the utility of type $e_2+\epsilon$ constant. In the two-outcome case, this new menu was derived from the old one by slightly decreasing $U_g(e)$ and increasing $U_b(e)$ for all types in $[e_2, e_2+\epsilon]$. 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 $e > e_2+\epsilon$ would not prefer to announce an effort $e \in [e_2, e_2+\epsilon]$. 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 $U'(y;e)$ such that $U'(y;e_2+\epsilon)$ is a mean-preserving decrease in "risk" for type $e_2+\epsilon$. (We put "risk" in quotes because we are comparing distributions of utilities, and not distributions of wage payments.) Formally,

$$\int_0^M U'(y;e_2+\epsilon)f(y;e_2+\epsilon) = \int_0^M U(y;e_2+\epsilon)f(y;e_2+\epsilon) \text{ and}$$

$$\int_0^y U'(y;e_2+\epsilon)f(y;e_2+\epsilon) < \int_0^y U(y;e_2+\epsilon)f(y;e_2+\epsilon) \text{ for all } 0 < y < M.$$

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.

6. Implementation with a Single Contract

We now 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.

This single contract cannot be riskless, because a riskless contract is interim-efficient and thus is not renegotiated, so a riskless contract induces effort $\underline{e}$ by the agent. This suggests that the optimal single contract should be highly risky, and be renegotiated towards more insurance at the interim stage. Theorem 6.1 below shows that it suffices to give the riskiest contract in the menu corresponding to the optimal renegotiation-proof contract.

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 6.1: The single *ex-ante* contract $(U_g^R(\hat{e}(R)), U_b^R(\hat{e}(R)))$ yields the same distribution of effort and payoffs as the optimal renegotiation-proof contract for rent R.

Remark: As in Theorem 5.1, we assume that if the agent's payoff is maximized by each of a range of actions E^* he is willing to use the probability distribution over E^* that the principal recommends. Section 4 showed how this mixed strategy can be "purified" if the principal offers an *ex-ante* contract menu and the agent has only two possible effort levels. We have not worked through the technical details involved in purifying a distribution that has a continuous density, and we do not know whether the purification argument extends to the case of a single *ex-ante* contract.

Proof of Theorem 6.1: 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 $\tilde{e}$, i.e., $\tilde{e} = \inf\{e | F(e) = 1\}$. 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:

$$(6.1)\quad \begin{aligned} &\text{(a)}\quad p(e)U_g(e)+(1-p(e))U_b(e) \geq p(e)U_g(e')+(1-p(e))U_b(e') \text{ for all } (e,e') \\ &\text{(b)}\quad p(e)U_g(e)+(1-p(e))U_b(e) \geq p(e)U_g^R(\hat{e}(R))+(1-p(e))U_b^R(\hat{e}(R)) \text{ for all } e. \end{aligned}$$

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

Let us show that there is no gap in the distribution; i.e., F(e) is strictly increasing for $e \leq \tilde{e}$. 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 not prefer to announce that it is e_2 must bind. If not,

$$(6.2)\qquad p(e_1)U_g(e_2)+(1-p(e_1))U_b(e_2) = p(e_1)U_g(e_1)+(1-p(e_1))U_b(e_1)-\Delta,$$

where $\Delta > 0$ from 6.1(a). Then the principal could increase his interim profit by offering the alternative contract

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

instead. Let us check that the new contract satisfies 6.1 (a) and (b). 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 $e' \le e_1$ because all utilities under e_1 are reduced uniformly. So we must prove that a type $e \le e_1$ does not want to choose the scheme of type $e' \ge e_2$. To see this, note that incentive compatibility of the original interim contract requires that the announcement-dependent utilities of type e_1 (which are also denoted $(U_g(e), U_b(e))$) satisfy $U_g(e) \ge U_b(e)$ and $U_g(e)-U_b(e)$ non-decreasing 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 $e \ge e_2$ 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

(6.3) $$p(e_2)U_g(e_2)+(1-p(e_2))U_b(e_2) \ge p(e_2)U_g(\hat{e}(R))+(1-p(e_2))U_b(\hat{e}(R)).$$

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 6.1 (b) for all $e \le e_1$.

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 (6.2). 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 $e \ge e_2$, 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 $e \leq e_0$, 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 $e < e_0$. Thus $e_0 = \underline{e}$.

The above arguments show that equilibrium distribution of efforts has support $[\underline{e},\tilde{e}]$ where $\tilde{e}_1 \leq +\infty$. The *ex-ante* equilibrium condition for the agent implies that for all e in $(\underline{e},\tilde{e}]$:

$$(6.3) \quad \begin{aligned} &\text{(a)} \quad p(e)U_g(e)+(1-p(e))U_b(e)-D(e) = R' \\ &\text{(b)} \quad p(e)\dot{U}_g(e)+(1-p(e))\dot{U}_b(e) = 0 \text{ a.e.} \end{aligned}$$

where $R' \geq R$ in the IR constraint 6.3 (a) (because the agent can always choose effort $\hat{e}(R)$ and refuse to renegotiate) and 6.3 (b) the choice of announcement is the first-order condition for the constraint that the given effort is optimal. (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 $[\underline{e},\tilde{e}]$, $(U_g(e), U_b(e)) = (U_g^{R'}(e), U_b^{R'}(e))$.

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 $\tilde{e}$:

$$(6.4) \qquad p(\tilde{e})U_g(\tilde{e})+(1-p(\tilde{e}))U_b(\tilde{e}) = p(\tilde{e})U_g^R(\hat{e}(R))+(1-p(\tilde{e}))U_b^R(\hat{e}(R)).$$

(If not, the principal would reduce the interim utility of all types 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 the proof of Theorem 5.1, 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, so that if the type who chooses $\tilde{e}$ is indifferent between announcing $\tilde{e}$ and announcing $\hat{e}(R)$, then it is strictly better to choose effort $\hat{e}(R)$ than effort $\tilde{e}$, and so $\tilde{e} = \hat{e}(R)$.

Finally, 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. As earlier, we now invoke the assumption that when the agent is indifferent, he is willing to play the mixed strategy that the principal recommends. Q.E.D.

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.

APPENDIX: PROOF OF THEOREM 5.1

Proof of Theorem 5.1: Fix a rent R. We first characterize incentive compatible allocations in lemmas A.1 and A.2. Lemmas A.3 and A.4 then compute the optimal renegotiation-proof contracts for a given R.

Lemma A.3 states the observation we have made previously that in the set of optimal efforts, choosing some effort and then the scheme associated with a different effort is a strictly inferior strategy.

Lemma A.1: Let $e_1 \neq e_2$, e_1 and e_2 belong to E^* and $e_2 > \underline{e}$. Then

$$\text{(A.1)} \qquad V(e_1) = p(e_1)U_g^R(e_1) + (1-p(e_1))U_b^R(e_1)$$

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

Proof of Lemma A.1: This follows from the fact that the contract $U^R(e_1)$ and $U^R(e_2)$ induce the agent to take effort e_1 and e_2 respectively.

Next, 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

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

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 $U_g(e) = \lim_{n\to\infty} U_g(\tilde{e}^n)$ and $U_b(e) = \lim_{n\to\infty} U_b(\tilde{e}^n)$. To simplify notation we use the similar notation for the announcement-dependent utilities $\{U_g^R(e), U_b^R(e)\}$ and for the utilities corresponding to a particular choice of effort $\{U_g(e), U_b(e)\}$. For effort levels in E* this is natural since by the revelation principle we consider only contracts where types report truthfully. The interim and *ex-post* utilities of types e not in E^* depends both on their actual effort level e and their choice of announcement $\tilde{e}$; we suppress this double dependence by substituting in the agent's optimal choice of announcement.

<u>Lemma A.2:</u> 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 $U_g(e) \geq U_b(e)$.

<u>Proof of Lemma A.2:</u> 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:

(A.2) (a) $V(e_1) \geq p(e_1)U_g(e_2) + (1-p(e_1))U_b(e_2)$

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

Using (A.5), we obtain

(A.3) $$[p(e_2)-p(e_1)][U_g(e_2)-U_b(e_2)] \geq$$

$$V(e_2)-V(e_1) \geq [p(e_2)-p(e_1)][U_g(e_1)-U_b(e_1)],$$

and taking $e_2 \to e_1$ in (A.6) yields the equation for $\dot{V}(e)$. Because $V(e)$ and $p(e)$ are increasing, (A.6) implies $U_g(e)-U_b(e)$ is non-negative 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.3 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.3: 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.3 Let us first show F(e) is strictly increasing. Suppose that there exists $e_1 < e_2$ such that $F(e_2+\epsilon)-F(e_2) > 0$ for all $\epsilon > 0$ 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 $[e_2,e_2+\epsilon]$ and keep the initial utilities for all other types, where δU_g and δU_b leave type $(e_2+\epsilon)$ with the same interim utility and satisfy $\delta U_b > 0$, so that all types in $[e_2,e_2+\epsilon]$ get more insurance in the new contract. We will show that the contract

$$\left\{(U_g(e),U_b(e))_{e \notin [e_2,e_2+\epsilon]},(U_g(e)+\delta U_g,U_b(e)+\delta U_b)_{e \in [e_2,e_2+\epsilon]}\right\}$$

(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 $[e_2, e_2+\epsilon]$, the equation for $\dot{V}(e)$ yields that the change in $\dot{V}(e)$ is negative. Since by construction, the change in $V(e_2+\epsilon)$ is zero, for e in $[e_2, e_2+\epsilon)$, 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 deviate to one of the new contracts. Incentive compatibility in the old contract implies that for all e and $\tilde{e}$ in $[e_2, e_2+\epsilon]$:

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

and hence

(A.5) $$p(e)(U_g(e)+\delta U_g) + (1-p(e))(U_b(e)+\delta U_b)$$

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

So, the new allocation is incentive compatible for types in $[e_2, e_2+\epsilon]$.

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 contracts. Lemma A.1, the fact that e_2 is *ex-ante* optimal for the agent, and the fact that E_1 is compact imply that for ϵ sufficiently small and $e \in E_1$

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

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

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

so that

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

Now, for δU_b sufficiently small, (A.6) also holds for the new allocation:

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

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

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

Because $\delta U_b > 0$ and $p(e) > p_2(e+\epsilon)$, $p(e)\delta U_g+(1-p(e))\delta U_b < 0$, and so (A.8) implies that

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

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

$$(A.9)\quad \delta\left[\int_{\underline{e}}^{+\infty}\pi(e)dF(e)\right] = -\int_{e_2}^{e_2+\epsilon}\left[p(e)\Phi'(U_g^R(e))\delta U_g+(1-p(e))\Phi'(U_b^R(e))\delta U_b\right]dF(e)$$

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

which, for ϵ small, is approximately

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

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).

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 exists 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

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

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

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

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 A.1. 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 $E^* \cap [\hat{e}(\delta U_b), e_2]$, let

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

and

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

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

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

$$- \int_{\hat{e}(\delta U_b)}^{\bar{e}_2} \left[p(e)\Phi'\left[U_g^R(e)\right]\delta U_g(e) + \left[1-p(e)\right]\Phi'\left[U_b^R(e)\right]\delta U_b(e)\right] dF(e).$$

The first term on the RHS of (A.11) is of order δU_b and strictly positive. The second term on the RHS 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.

We now determine the optimal renegotiation-proof distribution for the principal for a fixed rent R.

<u>Lemma A.4</u>: 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, $\hat{e}(R) > e^*(R)$.

<u>Proof of Lemma A.4</u>: From lemma A.3, 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

(A.12) $$V(e) = p(e)U_g(e)+(1-p(e))U_b(e).$$

As in the proof of lemma A.2, interim incentive compatibility requires that $V(\cdot)$ be continuous, increasing and a.e. differentiable in $[\underline{e},\hat{e}]$, with

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

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, as usual in mechanism design theory, 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

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

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

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

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

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

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

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

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

(A.18) $$\lambda \geq 0.$$

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

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

Therefore, from (A.15):

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

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:

(A.21) $$\frac{p(1-p)}{\dot{p}}\left[\Phi'(U_g^R)-\Phi'(U_b^R)\right]f$$

$$= \int_d^e\left[\left[p\Phi'(U_g^R)+(1-p)\Phi'(U_b^R)\right]dF - \lambda d\bar{e}\right].$$

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 not be tempted to renegotiate the contract of type e. The weight $f(e)de$ put on neighboring types could be shifted to higher efforts without jeopardizing renegotiation proofness. So intuition suggests that the optimum is obtained for $\lambda(\cdot) \equiv 0$.

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

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

The class of solutions of (A.22) is indexed by the value $f(\underline{e})$. That is, if $f(\cdot)$ is a solution, $\tilde{f}(e) \equiv \frac{\tilde{f}(\underline{e})}{f(\underline{e})} f(e)$ 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.

■ If $\dot{D}(e) = 0$, $\lim_{e \to \underline{e}^+} [\Phi'(U_g^R(e)) - \Phi'(U_b^R(e))] = 0$ from 5.1 (b) so that (A.22) 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,

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

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 $\underline{e}$.

■ 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 5.1 (b); 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

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

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

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

raises the principal's profit by

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

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 $\underset{-}{e}$ is

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

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

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

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

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

■ Thus, whatever $D(e)$, the optimal renegotiation-proof contract with rent zero belongs to the subclass of functions $\mathcal{F}$ indexed by $f(\underset{-}{e})$ and satisfying (A.22) (and therefore (A.30)). Fixing a $F(\cdot)$ (defined by $\underset{-}{\alpha}$ and $f(\cdot)$) belonging to $\mathcal{F}$, $F_\xi(\cdot) = \xi F(\cdot)$ also belongs to $\mathcal{F}$ for any ξ in $\left[0, \frac{1}{\underset{-}{\alpha}}\right]$, where $F(\cdot)$ corresponds to $\underset{-}{\alpha}_\xi = \zeta\underset{-}{\alpha}$ and $f_\xi(\cdot) = \xi f(\cdot)$. There is a one-to-one correspondence between the half-line $(\underset{-}{e}, +\infty)$ and $\mathcal{F}$: for any ζ there is an "upper bound" of F_ζ, $\hat{e}_\zeta$ in $(\underset{-}{e}, +\infty)$, defined by $\hat{e}_\xi = \inf\{e \mid F_\xi(e) = 1\}$; $\hat{e}_\xi$ is differentiable and strictly increasing in ξ, tends to $\underset{-}{e}$ when ξ tends to $1/\underset{-}{\alpha}$ 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 $\xi = 1/\underset{-}{\alpha}$) the full-insurance, no-incentive contract yielding $\pi^R(\underset{-}{e})$.

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

$$\text{(A.31)} \qquad \underset{\xi \in (0,1/\underline{\alpha})}{\text{Max}} \left\{ \xi\underline{\alpha}\pi^R(\underline{e}) + \int_{\underline{e}}^{\hat{e}_\xi} \pi^R(e)\xi f(e)de \right\}.$$

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:

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

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

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

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

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

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 ζ^* yields:

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

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

Note that this reasoning, which does not make use of $\lambda = 0$, 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 $\mathcal{F}$ (and thus 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 of G. We consider two cases.

<u>Case 1</u>: <u>$\hat{e} \leq e^*(R)$</u>. Let $F_\zeta(\cdot)$ 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_ζ 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_ζ 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:

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

(A.37) $$\dot{L}f_\zeta+L\dot{f}_\xi = Mf_\xi,$$

where

$$\text{(A.38)} \qquad L(e) = \frac{p(e)(1-p(e))}{\dot{p}(e)} [\Phi'(U_g^R(e)) - \Phi'(U_b^R(e))],$$

$$\text{(A.39)} \qquad M(e) = p(e)\Phi'(U_g^R(e)) + (1-p(e))\Phi'(U_b^R(e)).$$

Because $\lambda \geq 0$, (A.36) and (A.37) imply that $\dot{f}_\xi(e) \geq g(e)$ 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 $[\underline{e}, \hat{e}]$, $\int_{\underline{e}}^{\hat{e}} \pi^R(e)dF_\xi(e) \geq \int_{\underline{e}}^{\hat{e}} \pi^R(e)dG(e)$ with strict inequality unless $G = F_\xi$.

<u>Case 2</u>: <u>$\hat{e} > e^*(R)$</u>. 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 $\mathcal{F}$ 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

$$\text{(A.40)} \qquad H(e) = \begin{cases} F_\xi(e) & \text{for } e \leq e^*(R) \\ G(e) & \text{for } e \geq e^*(R) \end{cases}$$

The change in the principal's payoff is

$$\text{(A.41)} \qquad \int_{\underline{e}}^{\hat{e}} \pi^R(e)(dH(e) - dG(e)) = \int_{\underline{e}}^{e^*(R)} \pi^R(e)(dF_\xi(e) - dG(e)) \geq 0,$$

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 $G = F_\zeta$ a.e. up to $e^*(R)$; and that $G = F_\zeta$ 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 $\lambda \geq 0$:

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

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 ≥ e*(R):

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

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_ζ a.e. for e ≥ e*(R). Define the distribution K(·) with density k(·) by:

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

Because k(e*(R)) = g(e*(R)), (A.44) imply that k(e) ≥ g(e) for all e ≥ 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_ζ almost everywhere on $[\underline{e}, \hat{e}]$, so that the optimal solution belongs to $\mathcal{F}$ and satisfies (5.5). This concludes the proof of Theorem 5.1.

Q.E.D.

REFERENCES

Dewatripont, M. [1986], "Renegotiation and Information Revelation over Time in Optimal Labor Contracts", Ch.1 in On the Theory of Commitment with Applications to the Labor Market, Ph.D. Thesis, Harvard.

Fudenberg, D., Holmstrom, B. and P. Milgrom [1987], "Short-Term Contracts and Long-Term Agency Relationships", MIT mimeo.

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.

Hart, O. and J. Moore [1986], "Incomplete Contracts and Renegotiation", MIT mimeo.

Hart, O. and J. Tirole [1987], "Contract Renegotiation and Coasian Dynamics", forthcoming in Review of Economic Studies.

Holmstrom, B. [1979], "Moral Hazard and Observability", Bell Journal of Economics, 10, 74-91.

Holmstrom, B. and R. Myerson [1983], "Efficient and Durable Decision Rules with Incomplete Information", Econometrica, 51, 1799-1819.

Laffont, J.J. and J. Tirole [1988], "Adverse Selection and Renegotiation in Procurement", MIT mimeo.

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.

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, 407-430.

3-22-89