Econometrica, Vol. 60, No. 1 (January, 1992), 1–42

THE PRINCIPAL-AGENT RELATIONSHIP WITH AN INFORMED PRINCIPAL, II: COMMON VALUES[1]

BY ERIC MASKIN AND JEAN TIROLE

In many circumstances, a principal may have relevant private information when she proposes a contract to an agent. We analyze such a principal-agent relationship as a noncooperative game. The principal proposes a contract, which is accepted or rejected by the agent (who, for most of our analysis, has no private information). The contract is executed if accepted; otherwise, the reservation allocation takes effect. This allocation may be determined by a pre-existing contract (which the principal, by her proposal, is attempting to renegotiate), or it may simply be the no-trade point. In this paper, we assume that the principal's information directly affects the agent's payoff.

Before solving the game, we discuss Pareto efficiency with asymmetric information. We define an incentive-compatible allocation to be weakly interim efficient (WIE) if there exists no alternative incentive-compatible allocation that both parties prefer for all possible beliefs that the agent might have about the principal's private information (type). We show that any WIE allocation is interim-efficient (IE) for some beliefs.

The Rothschild-Stiglitz-Wilson (RSW) allocation relative to the reservation allocation $\mu_0^{\cdot}$ is the allocation that maximizes the payoff of each type of principal within the class of incentive-compatible allocations that guarantee the agent at least the utility he gets from $\mu_0^{\cdot}$ irrespective of his beliefs about the principal's type. The equilibrium set of the contract proposal game consists of the allocations that weakly Pareto dominate the RSW allocation. Thus, there is a unique equilibrium outcome if and only if the latter is IE (and the equilibrium outcome is the RSW allocation itself).

After characterizing the equilibrium allocations, we study those that are renegotiation-proof, when either the principal or the agent leads the renegotiation. We then compare our contract proposal game, which is a signaling model, with its "screening" counterpart. We conclude by extending our results to the case in which the agent as well as the principal has private information under the assumption of quasi-linear preferences.

KEYWORDS: Contract, principal-agent relationship, interim efficiency, signaling, renegotiation.

1. INTRODUCTION

STANDARD CONTRACT THEORY involving a principal and agent assumes that the principal, who proposes the contract, has no relevant private information at the (*ex-ante*) contracting date. This assumption is too restrictive in many economic circumstances, as the following examples illustrate: 1 (Public good). A government trying to elicit consumers' preferences for a public good has private information about the cost of supplying the good. 2 (Procurement). The Department of Defense has special knowledge about a weapon's strategic value when dealing with a defense contractor. 3 (Managerial Compensation). A manager has private information about her ability when bargaining over incentive contracts with an employer. 4 (Insurance). A shipping company seeking insurance against collision with icebergs knows the probability of collision. 5 (Franchising). A manufacturer offering a franchising agreement to a new retailer has private access to data about future demand for the product.

[1] This research was supported by the U.S. N.S.F., the Guggenheim Foundation, the British E.S.R.C., and St. John's College, Cambridge. We thank Bernard Caillaud, Jacques Crémer, Patrick Rey, and two referees for helpful comments.

When a party designing (or taking part in the design of) a contract has private information, the structure of the contract (and not just its execution) may reveal some of what he knows to other parties, as was emphasized by Myerson (1983).[2] Assume for simplicity that one party (the "principal") designs the contract and proposes it on a take-it-or-leave-it basis to another party (the "agent"). We shall use feminine pronouns for the principal and masculine ones for the agent. In general, the principal's private information can take two forms:

Private values: The principal's private information is not an argument of the agent's objective function (although the agent's private information, if any, may enter the principal's objective function).

Common values: The principal's private information is an argument of the agent's objective function.

The public good and procurement examples above belong to the private-values category; the consumers and the contractor do not care directly about the cost of supplying the public good or the strategic value of the weapon (although, indirectly, they may care about this information since it could affect the principal's behavior when the contract is executed). The compensation, insurance, and franchising models, by contrast, exemplify common values; the firm's payoff depends directly on the employee's ability (similarly, the insurance company's profit depends on the probability of paying collision benefits, and the retailer's revenue is affected by demand for the product).

It is important to note that many situations that would be labeled as private values in the absence of a prior contract exhibit common values if the status quo utilities result from some prior contract. For instance, the contractor may not care per se about the Department of Defense's utility for the weapon system. However, if the two are already bound by a previous contract specifying rewards for production and penalties for breach, then the contractor's reservation utility does depend on DoD's information, and so any contract *renegotiation* must be analyzed as a common-value situation.

Our earlier paper (Maskin-Tirole (1990a)) analyzed private values. In the model of that paper, the principal proposes a contract, which the agent accepts or refuses. If accepted, the contract is executed. Otherwise, the two parties do not transact. Hence, the model is a three-stage game: proposal, acceptance/refusal, execution. We assumed that both the principal and agent have private information in the first stage.

An important implication of the private-values assumption is that the principal can guarantee herself the same payoff she would get were her type (private information) known by the agent—the "full-information" case[3]—by proposing the full-information contract. Indeed, in the full-information case, the optimal contract consists simply of a menu of allocations from which the agent chooses. Private values ensure that, if confronted with this contract, he will make the

[2] Myerson's approach, however, differs markedly from our own. Whereas we are concerned mainly with characterizing the equilibria of a noncooperative game, he touches only briefly on noncooperative behavior. We discuss the relationship between his and our results in the final section.

[3] Actually, "full information" is a slight misnomer because the *agent* still has private information.

same choice and obtain the same utility whether or not he knows the principal's type. However, the principal can in general do even better than her full-information payoff by retaining some discretion at the contract execution stage.

In fact, the equilibrium outcomes of the three-stage game coincide with the Walrasian allocations of a fictitious economy in which the traders are the different types of principal and "exchange" the slack associated with the agent's individual rationality and incentive compatibility constraints. This characterization implies that equilibrium of the contract-proposal game exists, and is Pareto efficient (in a strong sense) and (locally) unique. Each type of principal is (generically, in the space of objective functions) strictly better off than under full information.

In this paper, we turn to common values. We retain the three-stage model of our previous paper but now assume that the principal's private information directly enters the agent's objective function. Conforming to most of the signaling literature, we assume that the agent has no private information (except in Section 8, where we consider two-sided uncertainty under the assumption that parties have quasi-linear preferences). Unlike her private-values counterpart, the principal in the common-values model may not be able to ensure her full-information payoff. In the Spence (1974) education model, for instance, a highly productive employee may be forced to invest in wasteful signaling activity (education) to avoid being mistaken for a less able employee. (Here, we are designating the employee as the contract proposer, i.e., as the principal). Our goal is to characterize the set of equilibrium contracts.

As in our earlier work, we define a contract very broadly: it is simply a game between the two parties, the outcomes of which are allocations. We endow the principal with all the bargaining power by having her propose the contract in the first period. The agent's expected payoff from the contract depends on his interim beliefs about the principal's type, where these beliefs are obtained by updating the agent's prior beliefs using the information conveyed by the contract proposal. Thus in the second stage, the agent accepts the contract if and only if his expected utility, given his interim beliefs, exceeds his expected reservation utility (his expected payoff if he refuses the contract). To be general, we allow reservation utility to be type contingent. Therefore its expectation must be computed using the interim beliefs about the principal's type.

If the agent refuses the principal's proposal, the game is over and players get their reservation utilities. If he accepts, the contract (which is itself a game) is then executed.[4]

In Section 3 we introduce the concept of weak interim Pareto efficiency. An allocation is weakly interim efficient (WIE) if it is incentive compatible (i.e., each type of principal prefers her allocation to that of any other type) and there exists no other such allocation that, regardless of the agent's beliefs about the principal's type, both parties prefer. A closely related concept is that of a Rothschild-Stiglitz-Wilson (RSW) allocation. An incentive-compatible alloca-

[4] Examples of contract proposal games with an informed principal and common values are found in Aghion-Bolton (1987), Gallini-Wright (1987), Gertner et al. (1988), and Stoughton-Talmor (1990).

tion is RSW relative to the reservation allocation $\mu_0^{\cdot}$ if it maximizes the payoff of each type of principal within the class of incentive-compatible allocations that ensure the agent at least the utility he gets from $\mu_0^{\cdot}$ no matter what his beliefs are. In Proposition 1 and its Corollary, we establish that any RSW allocation is WIE and that an WIE allocation is RSW relative to itself. We then show (Proposition 2) that RSW allocations have a simple structure when the principal's preferences satisfy a conventional sorting condition. The connection between WIE and ordinary interim efficiency is drawn in Propositions 3 and 4, where it is shown that any WIE allocation is interim efficient for some beliefs.

In Section 4 we present our main characterization result (Theorem 1), where we establish that the equilibrium set of our three-stage game consists of the incentive-compatible allocations that (weakly) Pareto dominate the RSW allocation relative to the reservation allocation. Therefore, the equilibrium of the contract proposal game is unique if and only if the RSW allocation is interim efficient. We apply Theorem 1 in Section 5 to contract negotiation when no prior contract binds the two parties (Propositions 6 and 7).

In Section 6 we turn to renegotiation. If the reservation allocation derives from a prior contract, that contract is weakly renegotiation-proof if there exists an equilibrium of the three-stage game in which it is not renegotiated. It is strongly renegotiation-proof if in no equilibrium is it renegotiated. Propositions 8 and 9 establish that weak and strong renegotiation-proofness correspond to WIE and interim efficiency respectively.

In Section 7, we compare the equilibrium set in the game of Sections 2 through 5, where the party with private information makes the contract proposal, to that where the uninformed party has (most of) this power. Proposition 12 shows that in the game where there are two uninformed parties who both propose a contract to a party with private information, there are many equilibria, including some that are Pareto dominated by the RSW allocation. When this game is modified to give some power to the informed party, however, the equilibrium set turns out to coincide with that of Theorem 1 (Proposition 13).

Section 8 extends most of our results to the case where the agent has private information, under the assumption that parties have quasi-linear preferences. Roughly speaking, our results extend to this case because, with quasi-linear preferences, the different types of principal do not gain by trading slack on the agent's individual rationality and incentive compatibility constraints (see our companion paper Maskin-Tirole (1990a)), and so the agent's private information creates no additional complication. Finally, Section 9 compares our results with those of Myerson (1983).

2. THE MODEL

A. *Objective Functions and Information*

There are two parties, a principal and an agent. The principal has a von Neumann-Morgenstern utility function $V^i(y, t)$, where y, a vector of observable and verifiable actions,[5] belongs to a compact, convex subset of $\mathbb{R}^n$; t is a monetary transfer (belonging to a compact interval) to the principal from the agent; and i denotes the principal's private information or type. The function V^i is continuously differentiable and concave in (y, t). The principal has a finite number of possible types $i = 1, \ldots, n$, with prior probabilities $\Pi = \{\Pi^i\}$ such that $\Sigma_{i=1}^n \Pi^i = 1$. We shall assume that a type i indifference curve is nowhere tangent to a type j indifference curve $(i \neq j)$.[6]

The agent has a von Neumann-Morgenstern utility function $U^i(y, t)$. The common-values assumption is embodied by making his utility depend on the principal's type i. We assume that $U^i(y, t)$ is strictly increasing in i for almost all (y, t) (a higher i corresponds to a "better" type). It also is continuously differentiable and concave in (y, t).

We assume compactness and regularity in order to ensure that optimal contracts exist and are well-behaved as functions of the parameters. Occasionally, however, we will drop the restriction of y and t to compact sets and invoke the following standard assumption (subscripts denote partial derivatives):

SORTING ASSUMPTION: (i) *y is one-dimensional, and y and t can be any real number*;

(ii) $U_y^i \geq 0$, *and there exists* $\varepsilon > 0$ *such that* $V_y^i < -\varepsilon$, $V_t^i > \varepsilon$, $U_t^i < -\varepsilon$;

(iii) *for all numbers* $\bar{u}$ *and* $\bar{v}$ *there exists a (finite) solution to the program* $\max V^i(y, t)$ *subject to* $\bar{v} \geq V^{i-1}(y, t)$ *and* $U^i(y, t) \geq \bar{u}$;

(iv) (*Sorting*) $(-V_y^i/V_t^i) > (-V_y^j/V_t^j)$ *for* $i < j$.

As we will see below, condition (iii), together with the other parts of the Sorting Assumption, ensures that optimal contracts exist, despite the lack of compactness. It, in turn, is implied by the assumption that along any indifference curve, the agent's marginal rate of substitution $-U_y^i/U_t^i$ goes to zero as $y \to \infty$, and goes to infinity as $y \to -\infty$.

The following examples satisfy the Sorting Assumption (except possibly for a change of domain of y and t):

Managerial Compensation: In this example, t refers to the manager's compensation and $y \geq 0$ to her performance[7] (e.g., output, profit, or cost reduction).

[5] Because these actions are observable and verifiable, it does not matter which party performs them. They correspond, however, to the signaling activity by the principal (informed party) in conventional signaling models.

[6] This no-tangency assumption is weaker than the Sorting Assumption below. Yet it is stronger than necessary. It is not required for Propositions 1 and 3. Proposition 5 needs only the weaker assumption that one can perturb the RSW allocation slightly so as to make the IR and IC constraints strictly binding. This weaker assumption is all that is needed for Propositions 5 through 13 and Theorem 1 (except for parts of Propositions 7 and 11).

[7] It may seem odd that, according to our terminology, the manager is designated the "principal" even though she is the party who performs. However, as noted in footnote 5, it does not actually matter in our framework "who does what." Hence, "principal" simply denotes the party who proposes the contract.

The utility functions are:

$$V^i(y,t) = t - C^i(y)$$

where for $i < j, C^i(y) > C^j(y)$ if $y \neq 0$ and $dC^i/dy > dC^j/dy > 0$, $d^2C^i/dy^2 > 0$; and

$$U^i(y,t) = iy - t.$$

The abler the manager (the higher the value of i), the lower is her marginal cost of performance and the higher is the employer's marginal profit. Letting $z \equiv C^i(y)$, we can rewrite the agent's utility function as $U^i(z,t) = i(C^i)^{-1}(z) - t$. If, for example, C^i satisfies $\lim_{z \to -\infty} d(C^i)^{-1}(z)/dz = \infty$, and $\lim_{z \to +\infty} d(C^i)^{-1}(z)/dz = 0$, condition (iii) of the Sorting Assumption is satisfied.

This compensation example embraces Spence's education model. However, in our setting, education is chosen *after* contracting (see the discussion of the principal-agent game below), which would make sense if, say, the employee's education or training were financed by the employer. We will at times refer to this example to illustrate our results.

Insurance: Consider a risk-averse shipper with von Neumann-Morgenstern utility function $W(\cdot)$. Her initial income is I. There are two states of nature: "iceberg collision" or "no collision." A collision entails a monetary loss L. The probability of no collision is α^i where $\alpha^i > \alpha^j$ for $i > j$. The shipper pays an insurance premium $-t$ and is reimbursed $-y$ in case of accident. Her utility function is thus: $V^i(y,t) = \alpha^i W(I+t) + (1-\alpha^i)W(I+t-L-y)$. A high-risk shipper ($\alpha^i$ small) is more eager to obtain insurance than a low-risk shipper. The agent (insurance company) is risk-neutral with objective function $U^i(y,t) = (1-\alpha^i)y - t$.

Franchising: A risk-neutral manufacturer offers a two-part tariff $t - yq$ to a risk-neutral retailer, where t is a franchise fee, y is the negative of the wholesale price, and q is the quantity sold. Consumer demand is $q = D(p,i) = a + i - p$ where p is the retail price chosen by the retailer, and i is the state of demand (known by the manufacturer). Assuming that the retailer chooses p before knowing i (this is not crucial), and ignoring distribution costs, we have: $p(y) = (a + Ei - y)/2$ and $q(y,i) = (a - Ei + y + 2i)/2$ (where $Ei = \sum_{i=1}^n \Pi^i i$). Letting c denote the manufacturing cost, we have:

$$V^i(y,t) = t - (y+c)q(y,i)$$

and

$$U^i(y,t) = (p(y)+y)D(p(y),i) - t.$$

As is easily checked, a high-demand manufacturer values a high wholesale price relatively more than a low-demand manufacturer.

Because incentive problems may be nonconvex, it may be desirable to allow for *random* outcomes. Accordingly, let $\mu(\cdot,\cdot)$ denote a probability measure on the cross product of the action and transfer sets. That is, roughly speaking, $\mu(y,t)$ represents the probability density of action y and transfer t. We will allow contracts to specify random outcomes μ. To simplify the notation, we define for all i in $\{1,\dots,n\}$

$$V^i(\mu) \equiv \int V^i(y,t)\, d\mu(y,t)$$

and

$$U^i(\mu) \equiv \int U^i(y,t)\, d\mu(y,t).$$

Note that $V^i(\cdot)$ and $U^i(\cdot)$ are linear in μ.

DEFINITION: An *allocation* is a menu $\mu^{\cdot} = \{\mu^i\}_{i=1}^n$ of (possibly random) outcomes, one for each type of principal.

DEFINITION: An allocation $\mu^{\cdot} = \{\mu^i\}_{i=1}^n$ is *incentive compatible* if, for all i and j, $V^i(\mu^i) \geqslant V^i(\mu^j)$.

DEFINITION: An allocation $\mu^{\cdot} = \{\mu^i\}_{i=1}^n$ *Pareto dominates* allocation $\bar{\mu}^{\cdot} = \{\bar{\mu}^i\}_{i=1}^n$ if $V^i(\mu^i) \geqslant V^i(\bar{\mu}^i)$ for all i, with strict inequality for some i.

Note that incentive compatibility and Pareto dominance are defined in reference to the principal's preferences only.

B. *The Principal-Agent Game*

Let us describe our three-stage game in detail. In the first stage the principal proposes a contract or mechanism in the feasible set M (we will use the words "contract" and "mechanism" interchangeably). A mechanism m in M specifies (i) a set of possible actions for each party and (ii) for each pair of moves s^p and s_a by the principal and agent, respectively, a corresponding measure $\mu(\cdot,\cdot)$ on the set of deterministic outcomes (y,t). Thus, a contract or mechanism is just a game form. The parties' actions can be thought of as announcements of payoff-irrelevant messages; the mechanism selects a (random) outcome conditional on these announcements. Observe that, because the principal can make announcements, she may be able to reveal information at the third stage (see below) as well as at the contract proposal stage. This fact will prove important in our analysis. In contrast, we will see that allowing the agent to make announcements does not affect the equilibrium set (when no prior contract is in effect—see Proposition 6), intuitively, because he has no private information to announce.[8]

[8] Nevertheless, moves by the agent *do* play a role in our treatment of renegotiation. Specifically, they may enable the agent to "punish" the principal should she ever propose a new contract (see Proposition 8*).

We suppose that M consists of all finite simultaneous-action mechanisms[9] (mechanisms where the players choose their actions simultaneously from finite sets). Notice that the set M includes the set of *direct revelation mechanisms*, (DRM's), in which the principal simply announces her type i, thus choosing from a "menu" (allocation) $\{\mu^i\}_{i=1}^n$. We will make considerable use of these DRM's by repeatedly invoking the revelation principle for Bayesian games.[10] In the present context this principle asserts that for any mechanism and for given beliefs at the time that mechanism is about to be played (i.e., after it has already been accepted), any equilibrium of the mechanism corresponds to a truthful equilibrium of a DRM.

In the second stage, the agent accepts or refuses the contract proposed by the principal. If he accepts, the two parties play the proposed mechanism in the third stage (for instance, the principal announces her type if the mechanism is a DRM), and the outcome corresponding to their third stage moves is implemented. The agent obtains his reservation utility if he rejects the contract. Thus, he will accept the proposed contract if and only if its expected utility exceeds the reservation level. The probabilities that he uses to compute expected utilities are the *interim beliefs* $\mathring{\Pi}^{\cdot} = \{\mathring{\Pi}^i\}_{i=1}^n$, obtained from the prior beliefs $\Pi^{\cdot}$ by updating on the basis of the principal's proposal.

A *perfect Bayesian equilibrium* is a vector of strategies[11]— one for the agent and one for each type of principal—and a vector of beliefs at each point in the game tree (formally, at each information set) such that (i) the strategies are optimal[12] at each point in the game tree (sequential rationality), (ii) interim beliefs $\mathring{\Pi}^{\cdot}(m)$ about the principal's type are the same at the beginning of the second stage as at the beginning of the third stage and are compatible with proposed offer m and the principal's presumed strategy (Bayesian updating), and (iii) given beliefs $\mathring{\Pi}^{\cdot}(m)$ at the beginning of the third stage, the third stage probability assessments are consistent in the sense of Kreps-Wilson (1982) (consistency). In other words, given beliefs $\mathring{\Pi}^{\cdot}(m)$, we require the continuation equilibrium beginning in stage 2 (after the principal has already proposed m) to be sequential.[13]

The technical reason for requiring the mechanisms in M to be finite is to ensure that a continuation equilibrium in the third stage exists. The simultaneous-action assumption guarantees that the continuation equilibrium correspondence as a function of interim beliefs $\mathring{\Pi}^{\cdot}$ is upper hemicontinuous (there can be

[9] We shall discuss the reasons for restricting M in this way below.

[10] Note that by appealing to DRM's we are not suggesting that they are "realistic." What one typically sees in actual contracts is a schedule in which compensation (t) is tied to output (y). This is of course equivalent to a DRM.

[11] These strategies for the overall game should not be confused with the equilibrium actions within the mechanism played at the third stage.

[12] I.e., each type is maximizing expected utility given beliefs and the other types' strategies.

[13] Moreover, our definition of an overall equilibrium is basically that of sequential equilibrium. However, the principal's strategy space in the first stage (the set of contracts) is necessarily infinite, and therefore consistency of beliefs between stages 1 and 2 is not well-defined. Thus we require only that Bayes' rule be used to obtain $\mathring{\Pi}^{\cdot}(m)$ when an equilibrium contract is proposed (condition (ii)). The reader can check that if the set of feasible contracts were restricted to a finite subset of M, our definition of equilibrium would coincide with that of sequential equilibrium.

failures of upper hemicontinuity at points where $\mathring{\Pi}^i = 0$ for some i if the principal moves before the agent in m). Without any changes in the formal arguments, we can expand M to include any other mechanisms for which existence and upper hemicontinuity hold.

As mentioned in the introduction, we will consider two different specifications of the agent's reservation utility:

No prior contract: In many applications, the agent's reservation utility is independent of the principal's type when no prior contract binds the two parties. However, our analysis can accommodate type-contingent reservation utility: $U_0^i = U^i(\mu_0^i)$, where μ_0^i is the exogenously given outcome that pertains should the agent reject the proposal by a principal of type i. One instance where reservation utilities are type-contingent is the franchising example, where the manufacturer can franchise a competing retailer if the agent turns down her proposal. In some applications, U_0^i may actually be *decreasing* in i. For example, a prospective licensee that turns down an exclusive licensing agreement for a process innovation could well be worse off the better the innovation if a rival then gets the license.

Renegotiation: In this case, it is supposed that the parties have signed an earlier contract that leads to allocation $\mu_0^{\cdot}$, and our contract proposal game can be thought of as a process of *renegotiation*. Because the allocation in which this process results might alternatively have been obtained by a more elaborate contract that is not renegotiated, our analysis of renegotiation will focus on characterizing the set of renegotiation-proof allocations.

In either case, we can assume without loss of generality that the reservation allocation $\mu_0^{\cdot}$ is incentive compatible.

To summarize, the principal's strategy in the three-stage game consists of a choice of mechanism and a choice of announcement (s^p) in that mechanism. The agent's strategy consists of the decision to accept or reject the mechanism and a choice of announcement (s_a) in the mechanism. Both of the agent's decisions are contingent on the mechanism proposed. We are interested in the perfect Bayesian equilibria of the overall game. Thus, in particular, we assume that the agent updates his beliefs about the principal's type using Bayes' rule after observing the contract she proposed. Similarly, we suppose that the principal revises her beliefs appropriately after observing that the agent has accepted the contract. In the continuation game of the third stage, there may, of course, be multiple equilibria. We suppose that the players can coordinate over these equilibria by means of some public randomizing device[14] such as a coin

[14] A randomizing device is "public" if its realizations are common knowledge. The technical reason for allowing public randomization is to ensure that the equilibrium payoff set of the continuation game is convex. We could alternatively allow (with no substantive change) players to use imperfect coordinating devices (different players observe different coin flips which are imperfectly correlated). Note that this randomization is in addition to that already built into the mechanism.

flip. If the coin turns up heads, they play one equilibrium; if tails, they play another. Thus, in the third stage, we permit (publicly) correlated equilibria.

Notice that in our model the values of both y and t are determined by the equilibrium contract. As noted earlier, this contrasts with some of the signaling literature (in particular, Spence (1974)), where y is chosen before contracting. Our approach follows in the tradition of the screening literature, where *all* variables are contractually set.

3. EFFICIENCY CONCEPTS

A. *Weakly Interim Efficient and RSW Allocations*

A.1. *General Definitions*

The following efficiency concepts play a crucial role in the rest of the paper. Notice that, although they will help us characterize the equilibria of our model, they are defined without reference to any game.

DEFINITION: An allocation $\bar{\mu}^{\cdot}=\{\bar{\mu}^i\}_{i=1}^n$ is *weakly interim efficient* (WIE) if (a) it is incentive compatible and (b) there exists no Pareto-dominating incentive compatible allocation $\mu^{\cdot}=\{\mu^i\}_{i=1}^n$ that, *regardless of the principal's type*, yields the agent at least as much utility. That is, $\bar{\mu}^{\cdot}$ is a solution to Program I[15] for some vector of positive weights $\{w^i\}_{i=1}^n$:

$$\text{Program I:} \quad \operatorname*{Max}_{\{\mu^{\cdot}\}} \sum_{i=1}^{n} w^i V^i(\mu^i) \qquad \text{subject to}$$

(IC) $\quad V^i(\mu^i) \geqslant V^i(\mu^j) \qquad$ for all i and j

and

(IR^i) $\quad U^i(\mu^i) \geqslant U^i(\bar{\mu}^i) \qquad$ for all i.

As we show in subsection B, weak interim efficiency is equivalent to ordinary interim efficiency relative to some beliefs.

To motivate the introduction of WIE, suppose that the principal and agent are initially bound by a contract in which the principal has the discretion to choose from an incentive compatible allocation $\bar{\mu}^{\cdot}=\{\bar{\mu}^i\}_{i=1}^n$. Then a *necessary* condition for all this allocation to be an equilibrium outcome of the three-stage proposal game—i.e., to be "renegotiation-proof"—is that the allocation be weakly interim efficient. Suppose instead that there exists a Pareto dominating allocation $\mu^{\cdot}$ satisfying (IC) and (IR^i) for all i. Let the principal propose the contract that allows her to choose from the allocation $\mu^{\cdot}=\{\mu^i\}_{i=1}^n$ (of course, if the proposal is rejected then she will choose from the allocation $\{\bar{\mu}^i\}_{i=1}^n$). The agent can accept the renegotiation offer $\mu^{\cdot}$ without risk, because (IC) and (IR^i) guarantee that, regardless of the principal's type, his utility will be at least as

[15] As Patrick Rey has pointed out to us, the set of WIE allocations gives rise to an $n-1$ dimensional, convex subset of the space of the utilities of the n types of principals.

large as before. Moreover, at least one type of principal is better off making this proposal, so that the initial contract $\bar{\mu}^{\cdot}$ is not renegotiation-proof. Weakly interim efficient allocations thus play an important role in the characterization of renegotiation-proof allocations.

DEFINITION: An allocation $\hat{\mu}^{\cdot}(\mu_0^{\cdot}) = \{\hat{\mu}^i(\mu_0^{\cdot})\}_{i=1}^n$ with associated payoffs $\hat{V}^{\cdot}(\mu_0^{\cdot}) = \{\hat{V}^i(\mu_0^{\cdot})\}_{i=1}^n$ is an *RSW* (*Rothschild-Stiglitz-Wilson*) *allocation*[16] *relative to the reservation allocation* $\mu_0^{\cdot}$ if and only if, for all i,

$$\text{Program II}^i: \qquad \hat{V}^i(\mu_0^{\cdot}) \equiv V^i(\hat{\mu}^i(\mu_0^{\cdot})) = \max_{\{\mu^{\cdot}\}} V^i(\mu^i) \qquad \text{subject to}$$

$$\text{(IC)} \qquad V^j(\mu^j) \geqslant V^j(\mu^l) \qquad \text{for all } j, l \in \{1, \dots, n\}$$

and

$$(\text{IR}_0^j) \qquad U^j(\mu^j) \geqslant U^j(\mu_0^j) \qquad \text{for all } j \in \{1, \dots, n\}.^{17}$$

That is, each type i maximizes her own utility within the set of allocations that are incentive compatible and, regardless of the principal's type, yields the agent at least his reservation utility.

It should be noted that an RSW allocation is defined by n independent optimizations, one for each type. From the continuity of the utility functions and the compactness of the domains, an RSW allocation exists. Clearly, the RSW payoffs $\hat{V}^{\cdot}(\mu_0^{\cdot})$ associated with a reservation allocation $\mu_0^{\cdot}$ are unique. The RSW allocation $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ associated with $\mu_0^{\cdot}$ need not, in general, be unique but turns out to be so in many well-known models (e.g., those of Spence (1974), Rothschild-Stiglitz (1976), and Wilson (1977)). Henceforth, for expositional convenience only, we will simply *assume* that $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ is a unique allocation (see footnote 19 for conditions sufficient for uniqueness). We will also assume that $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ is not on the boundary of the feasible set.[18]

To understand the significance of this concept, imagine that the principal and agent have not previously signed a contract. Let μ_0 denote the status-quo allocation if the principal's proposal is rejected. We claim that *the principal of type i can guarantee herself* $\hat{V}^i(\mu_0^{\cdot})$; she can simply propose the contract that gives her the discretion to choose from the menu that solves Program II^i after the agent has accepted. Because of (IC) and (IR_0^i), the agent does not suffer from accepting the proposal whatever his beliefs are about the principal's type. We thus conclude that *only allocations where, for all i, the type i principal's payoff is no less than* $\hat{V}^i(\mu_0^{\cdot})$ *are candidates for equilibrium of the contract proposal game*.

[16] We use the mnemonic term "RSW" allocation because, in the insurance model described above, this allocation is precisely the zero-profit separating allocation that figures prominently in Rothschild-Stiglitz (1976) and Wilson (1977).

[17] Notice that the set of allocations satisfying the constraints is nonempty since it includes $\mu_0^{\cdot}$ itself.

[18] This assumption is used in Proposition 5 only. If it is not satisfied, Proposition 5 still holds if neither of the two components (y, t) of the reservation outcome μ_0^i is on its respective boundary.

In some applications it is *not* the case that the type i principal's payoff is $V^i(\mu_0^i)$ if the agent rejects the proposal. For example, μ_0^i may correspond to an outcome that the agent achieves without the principal, in which case the principal gets $(0,0)$, not μ_0^i. Such a possibility creates no difficulty for our analysis, however, so long as $\hat{V}^i(\mu_0^{\cdot})$ is at least as great as type i's payoff in the absence of a contract. Henceforth, we shall, for convenience, assume that this inequality holds. Of course, it holds automatically if μ_0^i really is the type i's outcome should the agent reject.

In our effort to characterize the equilibria of this game, we first show that an RSW allocation is incentive compatible and weakly interim efficient.

PROPOSITION 1: *For any $\mu_0^{\cdot}$, $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ is weakly interim efficient (and thus incentive compatible).*

PROOF: Let $\{\mu_i^j\}_{j=1}^n$ denote a solution to II^i (so $\mu_i^i = \hat{\mu}^i(\mu_0^{\cdot})$). The constraint (IC) in the program requires that $V^j(\mu_i^j) \geqslant V^j(\hat{\mu}^i(\mu_0^{\cdot}))$. But the set of constraints in II^i is independent of i. Thus, $V^j(\hat{\mu}^j(\mu_0^{\cdot})) \geqslant V^j(\mu_i^j)$ and so $V^j(\hat{\mu}^j(\mu_0^{\cdot})) \geqslant V^j(\hat{\mu}^i(\mu_0^{\cdot}))$. Hence, $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ is incentive compatible. If it were not weakly interim efficient, there would exist an incentive compatible allocation $\mu^{\cdot}$ such that for some i, $V^i(\mu^i) > V^i(\hat{\mu}^i(\mu_0^{\cdot}))$ and for all j, $U^j(\mu^j) \geqslant U^j(\hat{\mu}^j(\mu_0^{\cdot}))$. But this contradicts the definition of $\hat{\mu}^i(\mu_0^{\cdot})$. *Q.E.D.*

COROLLARY: *If $\mu_0^{\cdot}$ is weakly interim efficient, then $\hat{\mu}^{\cdot}(\mu_0^{\cdot}) = \mu_0^{\cdot}$.*

PROOF: By assumption, $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ is unique. Moreover, $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ is weakly interim efficient from Proposition 1 and weakly Pareto dominates $\mu_0^{\cdot}$. *Q.E.D.*

A.2. *The Sorting Assumption and RSW Allocations*

Without imposing more structure on preferences, one cannot say which of the incentive constraints in the programs of subsection A.1 are binding. There are, however, many examples in the signaling literature where the RSW allocation relative to the reservation allocation has the property that only the "upward adjacent" constraints (the ones stipulating that the type i principal's utility be no less from her own allocation than from that of the type $i+1$) bind. As we shall see, this feature is a consequence of the examples' concentration on (deterministic) RSW allocations where $\mu_0^{\cdot}$ is the null allocation (i.e., $\mu_0^i = (0,0)$ for all i) and on their invoking a sorting condition. Indeed, when $U^i(\mu_0^i)$ is nonincreasing in i and the Sorting Assumption holds, the RSW allocation (restricted to be deterministic) relative to $\mu_0^{\cdot}$ is obtained by maximizing V^i subject only to $V^{i-1}(y^{i-1}, t^{i-1}) \geqslant V^{i-1}(y^i, t^i)$ and (IR_0^i) for all i:

PROPOSITION 2: *Suppose that the agent's reservation utility is nonincreasing in i ($U_0^1 \geqslant U_0^2 \geqslant \cdots \geqslant U_0^n$ where $U_0^i \equiv U^i(\mu_0^i)$) and that the Sorting Assumption* *holds. The RSW allocation (within the class of deterministic allocations)*[19] *relative to* μ_0 *is the "least-cost-separating allocation," obtained by successively solving the following programs*:

$$\textit{Program } III^1: \quad \operatorname*{Max}_{\{y^1, t^1\}} V^1(y^1, t^1) \quad \textit{subject to}$$

$$(IR_0^1) \quad U^1(y^1, t^1) = U_0^1$$

and for $k = 2, \dots, n$ *and given* $(y^1, t^1), \dots, (y^{k-1}, t^{k-1})$,

$$\textit{Program } III^k: \quad \operatorname*{Max}_{\{y^k, t^k\}} V^k(y^k, t^k) \quad \textit{such that}$$

$$(IC^k) \quad V^{k-1}(y^{k-1}, t^{k-1}) \geqslant V^{k-1}(y^k, t^k)$$

and

$$(IR_0^k) \quad U^k(y^k, t^k) \geqslant U_0^k.$$

In particular, the allocation satisfies constraint (IR_0^k) *with equality and* $y^{k-1} < y^k$ *and* $t^{k-1} < t^k$ *for all* k.

PROOF: See Appendix A.

REMARK 1: Although the individual rationality constraints in the solution to Program III^k hold with equality, the incentive compatibility constraints may not bind. While it is true that they *are* binding in the standard education and insurance models, this is because in those frameworks the signaling activity y is entirely wasteful (i.e., it does not directly raise the agent's payoff). See Subsection A.3 for an elaboration of this point.

REMARK 2: In incentive problems it is often helpful to impose only adjacent incentive compatibility constraints and then to deduce, as in Proposition 2, that a solution to the reduced program satisfies the full set of constraints. To make this deduction, however, it is ordinarily necessary to include the monotonicity constraints $y^i \leqslant y^{i+1}$ in the reduced program (see, for example, Mirrlees (1971)) or to impose a condition on the distribution of types (see Maskin-Riley (1984)). In Proposition 2, however, the monotonicity condition is *derived* rather than imposed. Monotonicity follows because, from the Programs III^k, the payoff of *each* type of principal is maximized individually, which is not the case in the Mirrlees or Maskin-Riley analyses.

[19] Here we simply restrict attention to deterministic allocations by assumption. It can be shown (see Maskin-Riley (1984)), however, that if $V^i(y,t)$ takes the form $t - \phi^i(y)$, where ϕ^i is strictly convex and $d^2\phi^i/dy^2$ is nondecreasing in i, then there is a unique deterministic solution to the Programs II^i. These conditions could be invoked in the compensation, insurance, and franchising examples presented above. (A change of variables such as the one we exhibited in the presentation of the compensation example is first required.) It is possible that Propositions 2 and 4 can be generalized to allow for a multidimensional action y, along the lines of Engers (1987) and Ramey (1988).

A.3. *The Compensation Example*

Let us illustrate our concepts in the compensation example of Section 2. Figures 1 and 2 represent the possible RSW allocations relative to the null allocation in the case of two types ($n = 2$). In Figure 1, the upward incentive

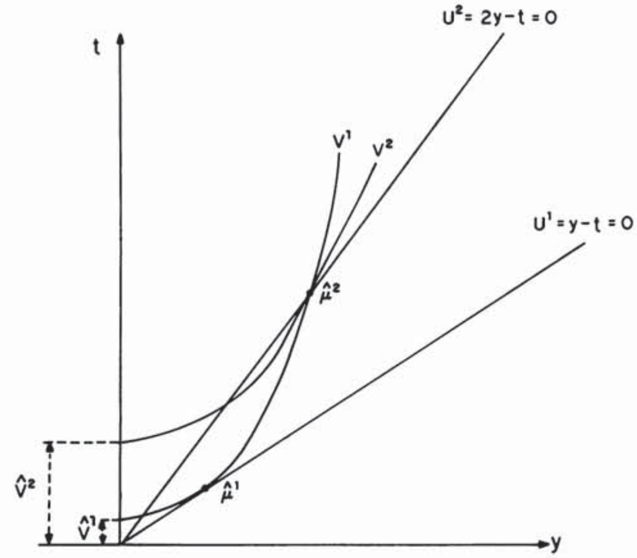

FIGURE 1

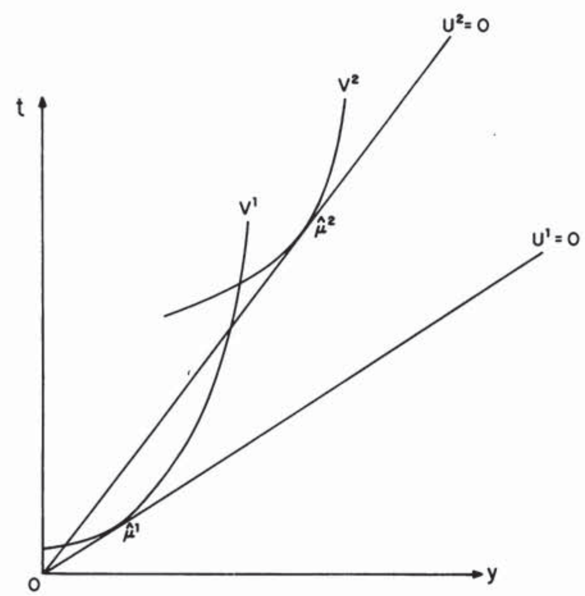

FIGURE 2

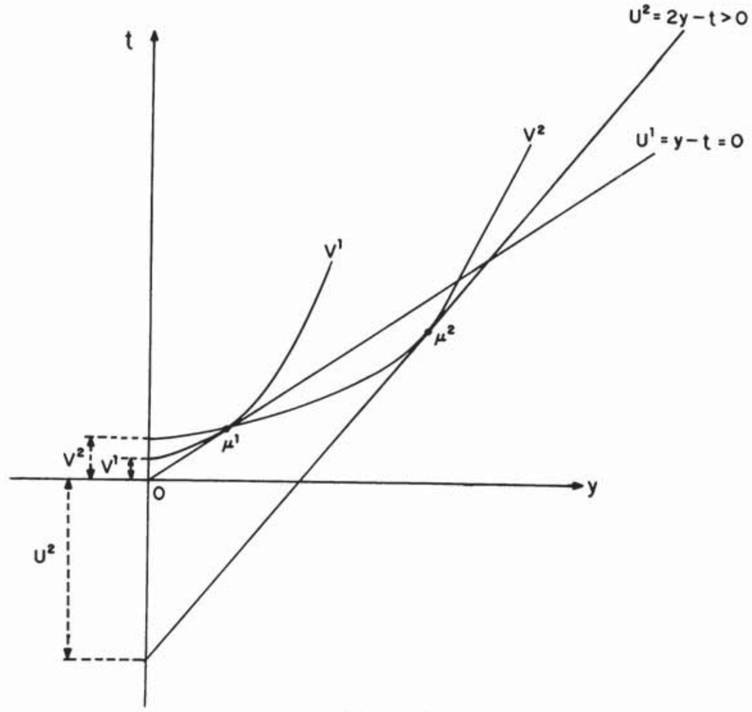

FIGURE 3

compatibility constraint is binding. The two straight lines correspond to the agent's isoprofit loci. The other curves are indifference loci for the two types of principal. Because the sorting condition ensures that the downward incentive compatibility constraint is satisfied automatically, $\hat{\mu}^1$ solves Program III[1] and is therefore (*ex-post*) efficient. Indeed, in Figure 1, it is the point of tangency between these type 1's indifference curve and the iso-profit curve $U^1(y,t)=0$. Because the upward constraint is binding, μ^2 is given by the intersection of $U^2(y,t)=0$ and $V^1(y,t)=\hat{V}^1$. This RSW allocation corresponds to the "least-cost-separating allocation" for type 2, which has received much emphasis in the work of Spence (1974), Rothschild-Stiglitz (1976), Riley (1979), and Cho-Kreps (1987), among others. It is clearly weakly interim efficient. To raise type 1's utility would require a decrease in $U^1(\mu^1)$. And, type 2's utility cannot be increased without violating either (IC) or (IR^1). Note also that the downward incentive constraint is not binding. Figure 1 is not the only possible RSW configuration. If the type-2 principal can signal sufficiently less expensively than the type 1, then, neither IC constraint is binding (i.e., $\hat{\mu}^2$ is *ex-post* efficient as well). This possibility is illustrated in Figure 2.

Figure 3 illustrates a weakly interim-efficient allocation for which the downward incentive compatibility constraint is binding. The outcomes μ^1 and μ^2 are *ex-post* efficient, and therefore $\mu^{\cdot}=\{\mu^1,\mu^2\}$ is weakly interim efficient. Type 2 is indifferent between μ^1 and μ^2, while type 1 strictly prefers μ^1. (Note that $U^2(\mu^2) > U^1(\mu^1)$ in this example.)

B. *Interim Efficient Allocations*

The efficiency concepts defined in subsection A require that the agent's individual rationality constraint be satisfied for each type of principal. That is, they are "belief-free." We now relax this requirement and demand only that allocations satisfy the agent's IR constraint on average, where the average is computed using the interim beliefs. We thus allow some individual (IR^i) constraints to be violated. This leads to the notion of interim efficiency in the sense of Holmström-Myerson (1983).[20]

DEFINITION: An allocation $\bar{\mu}^{\cdot} = \{\bar{\mu}^i\}_{i=1}^n$ is *interim efficient relative to beliefs* $\Pi^{\cdot}$ if (a) it is incentive compatible, and (b) there exists no other incentive compatible allocation $\mu^{\cdot} = \{\mu^i\}_{i=1}^n$ that Pareto dominates it, and yields the agent at least as much *expected* utility. Thus, $\bar{\mu}^{\cdot}$ is a solution to Program V for some vector of positive weights $\{w^i\}_{i=1}^n$:

$$\textit{Program V}: \qquad \operatorname*{Max}_{\{\mu^{\cdot}\}} \sum_{i=1}^{n} w^i V^i(\mu^i) \qquad \textit{such that}$$

$$(IC) \qquad V^i(\mu^i) \geqslant V^i(\mu^j) \qquad \textit{for all } i \textit{ and } j \textit{ and}$$

$$(IR) \qquad \sum_{i=1}^{n} \Pi^i U^i(\mu^i) \geqslant \sum_{i=1}^{n} \Pi^i U^i(\bar{\mu}^i).$$

C. *Relationship Between the Efficiency Concepts*

Because Program V is less constrained than Program I, an interim efficient allocation is weakly interim-efficient. One may wonder whether, for any weakly interim efficient allocation, there exist beliefs for which it is interim efficient. That this is indeed the case is shown in the following proposition.

PROPOSITION 3: *For any weakly interim efficient allocation $\bar{\mu}^{\cdot}$, the set of beliefs $\Pi(\bar{\mu}^{\cdot})$ such that this allocation is interim efficient is nonempty and convex.*

PROOF:[21] Let us first show that $\Pi(\bar{\mu}^{\cdot})$ is nonempty. Let C denote the set of incentive compatible contracts:

$$C = \left\{ \mu^{\cdot} = \{\mu^i\}_{i=1}^n \,|\, V^i(\mu^i) \geqslant V^i(\mu^j) \text{ for all } i \text{ and } j \right\}.$$

C is convex. A weakly interim-efficient allocation $\bar{\mu}^{\cdot}$ must satisfy $\bar{\mu}^{\cdot} \in$

[20] In this and the following subsection, we shall deal exclusively with interim beliefs and so, for notational simplicity, we will omit the "$\circ$" over "$\Pi^{\cdot}$."

[21] An earlier proof used a fixed-point method. This simpler proof was suggested to us by Jacques Crémer.

$\arg\max\{\sum_{i=1}^{n} w^i V^i(\mu^i) + \sum_i \nu^i (U^i(\mu^i) - U^i(\overline{\mu}^i))\}$ for some positive weights $\{w^i\}$ and nonnegative multipliers $\{\nu^i\}$. Let $\nu \equiv \sum_{i=1}^{n} \nu^i$. For each i, take $\Pi^i = \nu^i/\nu$ if $\nu > 0$ and $\Pi^i = 1/n$ otherwise. Note that $\sum_{i=1}^{n} \Pi^i = 1$. Thus, $\overline{\mu}^{\cdot}$ is weakly interim efficient if and only if there exist $\nu \geqslant 0$ and $(\Pi^1, \dots, \Pi^n)$ such that

$$\overline{\mu}^{\cdot} \in \underset{\mu^{\cdot} \in C}{\arg\max} \left\{ \sum_{i=1}^{n} w^i V^i(\mu^i) + \nu \sum_{i=1}^{n} \Pi^i \big(U^i(\mu^i) - U^i(\overline{\mu}^i)\big) \right\}.$$

Hence, $\overline{\mu}^{\cdot}$ is interim-efficient for beliefs $\Pi^{\cdot}$. (That the Lagrange conditions are necessary and sufficient results from the fact that the constraints are convex and admit an interior point.)

To see that $\Pi(\overline{\mu}^{\cdot})$ is convex, suppose that $\overline{\mu}^{\cdot}$ is interim efficient for beliefs $\Pi_1^{\cdot} = \{\Pi_1^i\}_{i=1}^n$ and $\Pi_2^{\cdot} = \{\Pi_2^i\}_{i=1}^n$. If it fails to be interim-efficient for beliefs $\lambda \Pi_1^{\cdot} + (1-\lambda)\Pi_2^{\cdot}$ $(0 < \lambda < 1)$, then it is Pareto-dominated by some incentive compatible allocation $\mu^{\cdot}$ for which

$$(*) \qquad \sum_i \big(\lambda \Pi_1^i + (1-\lambda)\Pi_2^i\big)\big(U^i(\mu^i) - U^i(\overline{\mu}^i)\big) \geqslant 0.$$

But because $\overline{\mu}^{\cdot}$ is interim efficient for beliefs $\Pi_1^{\cdot}$ and $\Pi_2^{\cdot}$,

$$\sum_i \Pi_1^i \big(U^i(\mu^i) - U^i(\overline{\mu}^i)\big) < 0,$$

and

$$\sum_i \Pi_2^i \big(U^i(\mu^i) - U^i(\overline{\mu}^i)\big) < 0,$$

which together contradict $(*)$. *Q.E.D.*

COROLLARY: *The set of beliefs for which an RSW allocation is interim efficient is nonempty and convex.*

In some circumstances, we can say more about the set of beliefs for which an RSW allocation is interim efficient:

PROPOSITION 4: *Adopt the hypotheses of Proposition 2. Suppose that* $\hat{\mu}^{\cdot} = \{(\hat{y}^1, \hat{t}^1), \dots, (\hat{y}^n, \hat{t}^n)\}$ *is a deterministic RSW allocation relative to* $\mu_0^{\cdot}$. *Then*, (a) *for any beliefs* $\Pi^{\cdot} \in \Pi(\hat{\mu}^{\cdot})$, $\Pi^i > 0$ *for all* i. *Moreover*: (b) *for all* k, *the suballocation* $\{(\hat{y}^1, \hat{t}^1), \dots, (\hat{y}^k, \hat{t}^k)\} \equiv {}^k\hat{\mu}^{\cdot}$ *is interim efficient in the submodel with types* $1, \dots, k$ *of principal for beliefs* $\{\alpha \Pi^1, \dots, \alpha \Pi^k\}$ *where* $\alpha = 1/(1 - \sum_{i=k+1}^{n} \Pi^i)$.

PROOF: See Appendix B.

The first part of Proposition 4 tells us that, when the Sorting Assumption holds and an RSW allocation is deterministic, the set of beliefs relative to which it is interim efficient consists entirely of strictly positive vectors. This result will come in handy below when we characterize the equilibria of the three-stage game (see Theorem 1). The second part establishes that, given the Sorting Assumption, if a deterministic RSW allocation is interim efficient with respect to beliefs $\Pi^{\cdot}$ then it remains interim efficient when we modify the model by deleting the top-most types of principal (and renormalizing the remaining probabilities appropriately).

4. EQUILIBRIUM CONTRACTS

We now characterize the equilibria of the contract proposal game. If the agent rejects the principal's offer, the allocation is $\mu_0^{\cdot}$. As mentioned earlier, $\mu_0^{\cdot}$ may be exogenously given (in the case where bargaining is over the initial contract) or result from a prior contract (in the case of renegotiation). Let $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ denote the RSW allocation relative to $\mu_0^{\cdot}$. The next proposition asserts that any type of principal gets at least her RSW payoff in equilibrium.

PROPOSITION 5: *In any equilibrium of the contract proposal game, the payoff of the type i principal is at least* $V^i(\hat{\mu}^i(\mu_0^{\cdot}))$.

PROOF: Choose $\varepsilon > 0$. Suppose that the type i principal proposes the "perturbed RSW allocation" $\mu^{\cdot} = \{\mu^j\}_{j=1}^n$ that solves Program II^i modified so that the right-hand side of (IR_0^j) is $U^j(\mu_0^j) + \varepsilon$ and (IC) is $V^j(\mu^j) \geqslant V^j(\mu^l) + \varepsilon$ for $l \neq j$. Because indifference curves for different types are never mutually tangent, these constraints can be satisfied if ε is sufficiently small.[22] Such a contract gives the principal the discretion to choose among the μ^j's, and the type j principal will surely choose μ^j. If the agent accepts the proposal, therefore he obtains more than his reservation utility whatever the principal's type. Hence the agent will accept the proposal. Since this is true for all ε, a lower bound on the type i principal utility is $V^i(\hat{\mu}^i(\mu_0^{\cdot}))$. *Q.E.D.*

REMARK: Proposition 5 illustrates a difference between our contract proposal game and the Spence (1974) education model: since, in the latter model, the screening variable (education) is chosen *before* contracting, the employee may not be able to guarantee herself the RSW payoff relative to the no-trade position. If we interpret y as education and t as wage in Figure 1, the RSW allocation is the least-cost-separating point $\{\hat{\mu}^1, \hat{\mu}^2\}$. But there are equilibria in which the employee chooses a level of education, and then proposes a wage such that her type 1 allocation is $\hat{\mu}^1$ and her type 2 allocation μ^2 is to the northeast of $\hat{\mu}^2$ on the locus $U^2 = 0$ (so that $V^2(\mu^2) < V^2(\hat{\mu}^2)$). The reason such points can be equilibria in Spence's model is that the type-2 employee could be mistaken for a type 1 if she chose the level of education corresponding to $\hat{\mu}^2$. In our framework, by contrast, the type 2 employee can include the option $\hat{\mu}^1$ in

[22] To see this, note that, for any interior pair (y, t) and ε small enough, the nontangency of indifference curves implies that we can find $\{(y^i, t^i)\}_{i=1}^n$ such that, for all i, (y^i, t^i) is within ε of (y, t) and $V^i(y^i, t^i) \geqslant V^i(y^j, t^j) + \varepsilon$ for all j.

the contract to guarantee that the firm does not suffer if it unluckily mistakes a low-productivity employee for a high-productivity one.

Our main result of this section establishes that the set of equilibrium allocations consists of incentive compatible allocations that weakly Pareto dominate the RSW allocation and are feasible for prior beliefs $\Pi^{\cdot}$.

THEOREM 1: *Suppose that the RSW allocation* $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ *is interim efficient relative to some beliefs* $\hat{\Pi}^{\cdot}$ *(not necessarily the prior beliefs), where* $\hat{\Pi}^i > 0$ *for all* i. *The set of equilibrium allocations of the contract proposal game (with respect to prior beliefs* $\Pi^{\cdot}$*) is the set of allocations* $\mu^{\cdot} = \{\mu^i\}_{i=1}^n$ *satisfying*

$$(IC) \qquad V^i(\mu^i) \geqslant V^i(\mu^j) \qquad \textit{for all } i, j,$$

$$(IR) \qquad \sum_i \Pi^i U^i(\mu^i) \geqslant \sum_i \Pi^i U^i(\mu_0^i),$$

and

$$V^i(\mu^i) \geqslant V^i(\hat{\mu}^i(\mu_0^{\cdot})) \qquad \textit{for all } i.$$

COROLLARY: *Under the hypothesis of Theorem* 1, *the equilibrium payoffs of the contract proposal game are unique if and only if the RSW allocation relative to* $\mu_0^{\cdot}$ *is interim efficient for the prior beliefs* $\Pi^{\cdot}$.

REMARK 1: Propositions 1 and 3 imply that the RSW allocation is interim efficient with respect to some beliefs. The substantive hypothesis in Theorem 1 is thus that it is interim efficient relative to some *strictly positive* beliefs.

REMARK 2: When $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ is interim efficient relative to $\Pi^{\cdot}$, the equilibrium of our three-stage game is also interim efficient (since the equilibrium outcome is $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ itself). When $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ is not interim efficient, however, there are in general many equilibria that are not IE. Indeed, there are many that are not even WIE (even though they dominate the RSW allocation, which is WIE). This lack of efficiency contrasts with the strong optimality results that obtain with private values (Maskin-Tirole (1990a)).

REMARK 3: Thanks to Proposition 4, the hypothesis of Theorem 1 is automatically satisfied if preferences satisfy the Sorting Assumption, if there exists a deterministic RSW allocation relative to $\mu_0^{\cdot}$, and if U_0^i is nonincreasing in i.

PROOF OF THEOREM 1: From Proposition 5, each type i can guarantee herself $\hat{V}^i = V^i(\hat{\mu}^i(\mu_0^{\cdot}))$. Hence, only allocations that Pareto dominate the RSW allocation relative to $\mu_0^{\cdot}$ are candidates for equilibrium.

Choose an incentive compatible allocation $\bar{\mu}^{\cdot} = \{\bar{\mu}^1, \dots, \bar{\mu}^n\}$ satisfying:

$$V^i(\bar{\mu}^i) \geqslant \hat{V}^i \qquad \text{for all } i$$

and

$$\sum_i \Pi^i U^i(\overline{\mu}^i) \geqslant \sum_i \Pi^i U^i(\mu_0^i).$$

Consider the following candidate equilibrium strategies: "The principal, whatever her type, proposes the contract $\overline{\mu}^{\cdot}$, giving her the discretion to choose within the set $\{\overline{\mu}^i\}$ after the contract is signed. The agent accepts this contract. If the principal proposes an alternative contract (mechanism) m, the agent's interim beliefs are $\mathring{\Pi}^{\cdot}$, and the associated equilibrium payoffs are $\mathring{V}^{\cdot} = (\mathring{V}^1, \ldots, \mathring{V}^n)$, where $\mathring{V}^i \leqslant \hat{V}^i$ for all i, and where $\mathring{\Pi}^{\cdot}$ and $\mathring{V}^{\cdot}$ depend on m and are specified as below." Clearly, if for all $m \neq \overline{\mu}^{\cdot}$, we can find such $\mathring{\Pi}^{\cdot}$ and $\mathring{V}^{\cdot}$, the principal will not gain by deviating from $\overline{\mu}^{\cdot}$. Furthermore, since the interim beliefs corresponding to the proposal $\overline{\mu}^{\cdot}$ are just the prior beliefs, the agent cannot profit by rejecting the proposal $\overline{\mu}^{\cdot}$.

Thus, it remains only to show that, for an arbitrary $m \neq \overline{\mu}^{\cdot}$, there exist out-of-equilibrium beliefs $\mathring{\Pi}^{\cdot}$ and an associated equilibrium of the continuation game (beginning in the second stage) defined by m in which no type of principal is better off than in the RSW allocation.

Suppose, to the contrary, that there exists m for which, for any beliefs $\mathring{\Pi}^{\cdot}$ and any corresponding equilibrium payoffs $\mathring{V}^{\cdot}, \mathring{V}^i > \hat{V}^i$ for some i. Now, for any $\mathring{\Pi}^{\cdot}$, let $\psi_m(\mathring{\Pi}^{\cdot}) = \{\mathring{V}^{\cdot} | \mathring{V}^{\cdot}$ is an equilibrium payoff vector for m when beliefs are $\mathring{\Pi}^{\cdot}\}$.

From our choice of the space of mechanisms, we have ensured that ψ_m is upper hemicontinuous and nonempty-valued (condition (iii) of our equilibrium definition states that the continuation equilibrium after m is proposed is a sequential equilibrium for any interim beliefs $\mathring{\Pi}^{\cdot}$). Because players can avail themselves of public randomizing devices, it is also convex-valued.

Fix $\delta \in (0,1]$. For any $\mathring{V}^{\cdot}$, let

$$\rho_\delta^i(\mathring{V}^{\cdot}) = \left\{ \tilde{p}^i | \tilde{p}^i \in \arg\max \left[p^i \mathring{V}^i + (1 - p^i)\hat{V}^i \right] \text{ subject to } p^i \in [\delta, 1] \right\},$$

and

$$\rho_\delta^{\cdot}(\mathring{V}^{\cdot}) = \rho_\delta^1(\mathring{V}^{\cdot}) \times \cdots \times \rho_\delta^n(\mathring{V}^{\cdot}).$$

That is, if the type i principal is offered a choice between $\mathring{V}^i$ and $\hat{V}^i$, $\rho_\delta^i(\mathring{V}^{\cdot})$ is the payoff-maximizing probability of choosing $\mathring{V}^i$, subject to the constraint that it be at least δ.

For any vector $p^{\cdot} \in [\delta, 1]^n$, let

$$\phi^i(p^{\cdot}) = p^i \hat{\Pi}^i \Bigg/ \sum_{j=1}^n p^j \hat{\Pi}^j.$$

That is, given that the principal has chosen $\mathring{V}^{\cdot}$ over $\hat{V}^{\cdot}$, $\phi^i(p^{\cdot})$ is the conditional probability that her type is i if the vector of prior probabilities is $\hat{\Pi}^{\cdot}$. Take $\phi^{\cdot}(p^{\cdot}) = (\phi^1(p^{\cdot}), \ldots, \phi^n(p^{\cdot}))$.

Let $\mathscr{V}_m$ be a compact and convex set containing all the equilibrium payoffs $\mathring{V}^{\cdot}$ of m for each $\mathring{\Pi}^{\cdot}$. Consider the correspondence from $[\delta, 1]^n \times \Delta^{n-1} \times \mathscr{V}_m$ to itself that maps $(p^{\cdot}, \Pi^{\cdot}, V^{\cdot})$ to $\rho_\delta^{\cdot}(V^{\cdot}) \times \phi^{\cdot}(p^{\cdot}) \times \psi_m(\Pi^{\cdot})$. This correspondence is upper hemicontinuous and convex-valued. Hence it has a fixed point $(p_\delta^{\cdot}, \Pi_\delta^{\cdot}, V_\delta^{\cdot})$. Let $(p_*^{\cdot}, \Pi_*^{\cdot}, V_*^{\cdot}) = \lim_{\delta \to 0}(p_\delta^{\cdot}, \Pi_\delta^{\cdot}, V_\delta^{\cdot})$. (There always exists a convergent subsequence. Moreover, $\phi^{\cdot}(p_*^{\cdot})$ is well-defined because $V_*^i > \hat{V}^i$ for some i. Hence, $p_*^i > 0$ and so $\Sigma_{j=1}^n p_*^j \hat{\Pi}^j > 0$.) Then $V_*^{\cdot} \in \psi_m(\Pi_*^{\cdot})$, $p_*^{\cdot} \in \rho_0(V_*^{\cdot})$, and $\Pi_*^{\cdot} \in \phi^{\cdot}(p_*^{\cdot})$.

Now let $\mu_*^{\cdot}$ denote the allocation corresponding to $V_*^{\cdot}$ and let $\mu_{**}^{\cdot}$ be defined by $\mu_{**}^i = p_*^i \mu_*^i + (1 - p_*^i)\hat{\mu}^i(\mu_0^{\cdot})$ (that is, each type i gets the outcome she prefers between μ_*^i and $\hat{\mu}^i(\mu_0^{\cdot})$). Then $\mu_{**}^{\cdot}$ is incentive compatible, Pareto dominates $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$, and satisfies $\Sigma_i \hat{\Pi}^i U^i(\mu_{**}^i) \geqslant \Sigma_i \hat{\Pi}^i U^i(\hat{\mu}^i(\mu_0^{\cdot}))$ (the last inequality holds because $\mu_{**}^{\cdot}$ is defined with reference to prior beliefs $\hat{\Pi}^{\cdot}$ and because the agent always has the option of refusing the proposal), contradicting the interim efficiency of $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ relative to $\hat{\Pi}^{\cdot}$ *Q.E.D.*

5. APPLICATION 1: INITIAL CONTRACTS

A. *Equilibrium Contracts*

If no previous contract is in force at the time of negotiation, we simply interpret $\mu_0^{\cdot}$ to be the allocation that arises if the principal and agent fail to sign a contract. Hence, from Theorem 1 we immediately obtain the following proposition.

PROPOSITION 6: *If no prior contract is in force, the set of equilibrium allocations of the contract proposal game coincides with the set of incentive-compatible allocations that weakly Pareto dominate the RSW allocation associated with the reservation allocation and are individually rational for the agent with prior beliefs $\Pi^{\cdot}$. Thus, the equilibrium is unique if and only if $\Pi^{\cdot}$ belongs to the convex set $\Pi(\hat{\mu}^{\cdot}(\mu_0^{\cdot}))$.*[23]

Let us illustrate Proposition 6 using the compensation example and assuming that $\mu_0^{\cdot}$ is the null allocation. Let Π^1 and Π^2 denote the prior probabilities of the two types. There exists $\hat{\Pi}^2 \in (0, 1]$ such that if $0 < \Pi^2 \leqslant \hat{\Pi}^2$, the unique equilibrium of the contract design game is the least-cost-separating allocation $\{\hat{\mu}^1, \hat{\mu}^2\}$ depicted in Figure 1. If $\Pi^2 > \hat{\Pi}^2$, there exists a continuum of equilibria, which all Pareto-dominate the least-cost-separating allocation. They all have the property that the agent loses relative to the null allocation if the principal is of type 1, but strictly gains if the principal is of type 2.

[23] Throughout this section and the next, we will assume that this set includes a strictly positive vector (as will be true, for example, under the hypotheses of Proposition 4).

B. *Refinements*

When the RSW allocation associated with the reservation allocation is not interim efficient, the theory implicitly makes a continuum of alternative predictions about the outcome. However, one may feel (although we ourselves are agnostic about this) that some of these outcomes are more reasonable than others and so advocate the use of an equilibrium refinement. In this section and the next, we will focus on two widely used refinements for extensive form games, those of Cho-Kreps (1987) (CK) and Farrell (1985)/Grossman-Perry (1986) (FGP).

In our context, the CK and FGP criteria take the following forms. (The definitions are somewhat more elaborate than usual because we deal with a three-stage, rather than a standard two-stage signaling game.) Let $T = \{1, \dots, n\}$ denote the set of all types. Consider a candidate equilibrium allocation $\mu^{\cdot}_{*} = \{\mu^{i}_{*}\}_{i \in T}$, yielding utilities $V^i_* = V^i(\mu^i_*)$ and $U^i_* = U^i(\mu^i_*)$ to the principal and the agent, when the principal is of type i. Let $\Pi^{\cdot} = \{\Pi^i\}_{i \in T}$ denote the agent's beliefs about the principal at the beginning of the contract proposal game. Let m denote an out-of-equilibrium mechanism offered by the principal (m corresponds to an off-the-equilibrium-path message sent by the sender in the traditional signaling game). $\mathring{\Pi}^{\cdot} = \{\mathring{\Pi}^i\}_{i \in T}$ will denote the interim beliefs following the contract offer m. Let $BR(\mathring{\Pi}^{\cdot}, m)$ denote the *equilibrium* allocations of the continuation game between the principal and the agent after m has been offered and has led the agent to update his beliefs to $\mathring{\Pi}^{\cdot}$. (We use the notation $BR(\mathring{\Pi}^{\cdot}, m)$ because, in the traditional signaling model, this set simply consists of the receiver's best responses to the message m given his beliefs $\mathring{\Pi}^{\cdot}$. In our framework, by contrast, m is itself a game, and so $BR(\mathring{\Pi}^{\cdot}, m)$ consists of continuation *equilibria*. If, however, we restricted the principal to propose direct revelation mechanisms, then $BR(\mathring{\Pi}^{\cdot}, m)$ would indeed be the set of best responses by the agent. The reader can check that Propositions 7 and 11 would be unaffected by such a restriction.) Let Δ^S denote the set of beliefs concentrated on a subset S of T: $\Delta^S \equiv \{\mathring{\Pi}^{\cdot} \mid \mathring{\Pi}^i = 0 \text{ if } i \notin S\}$. Let $BR(S, m) \equiv \bigcup BR(\mathring{\Pi}, m)_{\{\mathring{\Pi}^{\cdot} \in \text{int}\, \Delta^S\}}$. That is, $BR(S, m)$ is the set of possible equilibrium allocations of m when the agent puts weight on all types in S and no weight on any other types.

The equilibrium allocation $\mu^{\cdot}_{*}$ fails to pass the CK intuitive criterion if and only if there exists a mechanism m and a subset J of types (possibly empty) such that

(1) $\forall i \in J,\ \forall \mu^{\cdot} \in BR(T, m), \quad V^i_* > V^i(\mu^i)$ and

(2) letting $S = T \setminus J$, $\forall \mu^{\cdot} \in BR(S, m)$, there exists $i \in S$ such that $V^i_* < V^i(\mu^i)$.

In words, a type in J would lose by deviating and proposing contract m; and, given that beliefs put all weight on the complementary subset S, at least one type i strictly gains from deviating for any given continuation equilibrium.

Turning to the FGP criterion, consider a candidate equilibrium allocation $\mu^{\cdot}_{*}$, and alternative allocation $\mu^{\cdot}$ and a mechanism m. Let $\Gamma(\mu^{\cdot}, \mu^{\cdot}_{*}) = \{\mathring{\Pi}^{\cdot} \mid \mathring{\Pi}^i = 0 \text{ if } V^i(\mu^i) < V^i_*, \text{ and } \mathring{\Pi}^i/\Pi^i \leqslant \mathring{\Pi}^j/\Pi^j \text{ for all } i \text{ and } j \text{ such that } V^j(\mu^j) > V^j_*\}$.

That is, the agent's interim beliefs preserve the relative probabilities for the types who strictly gain in the allocation $\mu^{\cdot}$. The interim probability of a type who strictly loses is 0. And the ratio of the interim to the prior probability of a type who is indifferent is no greater than that for one who strictly gains. The allocation $\mu^{\cdot}_{*}$ fails to pass the FGP criterion if and only if there exist $m, \mu^{\cdot}$, and i such that $\mu^{\cdot} \in BR(\Gamma(\mu^{\cdot}, \mu^{\cdot}_{*}), m) \equiv \bigcup BR(\mathring{\Pi}^{\cdot}, m)_{\{\mathring{\Pi}^{\cdot} \in \Gamma(\mu^{\cdot}, \mu^{\cdot}_{*})\}}$, and $V^{i}(\mu^{i}) > V^{i}(\mu^{i}_{*})$. In words, for some interim beliefs $\mathring{\Pi}^{\cdot}$ in $\Gamma(\mu^{\cdot}, \mu^{\cdot}_{*})$, $\mu^{\cdot}$ is an equilibrium allocation of mechanism m in which all types i for which $\mathring{\Pi}^{i} > 0$ do at least as well under $\mu^{\cdot}$ as $\mu^{\cdot}_{*}$ (and some type does strictly better under $\mu^{\cdot}$). This set of types can be thought of as the "deviating coalition."

Proposition 7: (CK) *Suppose that there exist strictly positive beliefs* $\hat{\Pi}^{\cdot} \in \Pi(\hat{\mu}^{\cdot}(\mu^{\cdot}_{0}))$. *The RSW allocation* $\hat{\mu}^{\cdot}(\mu^{\cdot}_{0})$ *passes the CK requirement. It is the unique such allocation if* $\Pi^{\cdot} \in \Pi(\hat{\mu}^{\cdot}(\mu^{\cdot}_{0}))$. *Moreover, even if* $\Pi^{\cdot} \notin \Pi(\hat{\mu}^{\cdot}(\mu^{\cdot}_{0}))$, $\hat{\mu}^{\cdot}(\mu^{\cdot}_{0})$ *is the unique allocation provided that either* (*i*) $n = 2$ *and we ignore allocations on the boundary of the feasible set or both* (*ii*) *the Sorting Assumption holds and* (*iii*) *reservation utilities are nonincreasing* $(U_0^1 \geqslant U_0^2 \geqslant \ldots \geqslant U_0^n)$.

(FGP) *If* $\Pi^{\cdot} \in \Pi(\hat{\mu}^{\cdot}(\mu^{\cdot}_{0}))$ *and if* (*iv*) *for all beliefs* $\tilde{\Pi}^{\cdot}$, *any IE allocation* $\mu^{\cdot}$ (*relative to* $\tilde{\Pi}^{\cdot}$) *that weakly Pareto dominates* $\hat{\mu}^{\cdot}(\mu^{\cdot}_{0})$ *satisfies* $\sum_i \tilde{\Pi}^{i}(U^{i}(\mu^{i}) - U_0^{i}) = 0$, *then* $\hat{\mu}^{\cdot}(\mu^{\cdot}_{0})$ *passes the FGP criterion. If* $\Pi^{\cdot} \notin \Pi(\hat{\mu}^{\cdot}(\mu^{\cdot}_{0}))$ *and either* (*i*) *or* (*ii*) *above holds, then there exists no allocation passing the FGP requirement.*

Remark: Observe that, if the Sorting Assumption (including the unboundedness condition in part (i)) is satisfied, then any allocation is automatically interior and (iv) holds. (Since in some applications y and t may not be unbounded in both directions, one may have corner solutions.) Hence, in this case, we can conclude that (a) the RSW allocation passes the FGP criterion if and only if it is IE and (b) no other allocation passes the FGP criterion. Note too, that, when $n \geqslant 3$ and the RSW allocation is not interim efficient we require both (ii) and (iii) to show that any CK allocation must be RSW, whereas only the former condition is required to show that no FGP allocation exists.

The proof of Proposition 7 is in Maskin-Tirole (1990b). The result for the CK criterion selection has a familiar flavor thanks to the work of Cho-Kreps and Cho-Sobel (1987), and the same is true for the logic behind it (see Section 6B for the intuition). Note, however, that our framework differs from the earlier papers in that the sender's message is a contract proposal (rather than a one-dimensional signal), and reservation utilities may be decreasing (rather than constant).

6. APPLICATION 2: RENEGOTIATION

A. *Renegotiation-proof Allocations*

We now assume that the reservation allocation $\mu^{\cdot}_{0}$ corresponds to a previous contract. More precisely, we suppose that the two parties are initially bound by a contract that specifies that the principal can choose from the menu $\{\mu_0^1, \ldots, \mu_0^n\}$, i.e., the contract is a direct revelation mechanism. (We discuss below how our results are changed when more general initial contracts are allowed.) The principal's contract proposal is, therefore, an offer to renegotiate.[24] In this context, the agent's priors are his beliefs about the principal at the time of renegotiation.

Even in a world where parties can costlessly sign and enforce long-term contracts that are contingent on all observables, the contract they would sign if they could commit themselves not to renegotiate is generally time-inconsistent. That is, at some stage in the execution of this full-commitment contract, the parties might mutually benefit from renegotiation. Because any renegotiation can alternatively be built into the initial contract itself, characterizing the set of allocations that can arise when renegotiation is possible amounts to characterizing the set of *renegotiation-proof* contracts.[25]

When only one party has private information at the renegotiation stage, it is often assumed that the uninformed party (in our model, the agent) proposes the new contract. This assumption ensures that the contract proposal itself does not reveal information. Naturally, giving full bargaining power to the uninformed party at the renegotiation stage is extreme. One wishes to know how alternative distributions of power (specifically, permitting the informed party to propose the new contract) affect the outcome of the overall agency relationship. We now examine this question in our principal-agent framework.

Assume that the two parties are bound by a prior allocation $\mu_0^{\cdot}$. Suppose, for now, that the principal (the informed party) proposes a new contract m. The agent either accepts or rejects m. In the latter case $\mu_0^{\cdot}$ remains in force.[26] In the former case, $\mu_0^{\cdot}$ is supplanted and the two parties play mechanism m. We wish to characterize the allocations $\mu_0^{\cdot}$ that are renegotiation-proof (i.e., those that will not be supplanted in this renegotiation game). However, Theorem 1 implies that the outcome of the renegotiation game for an initial allocation $\mu_0^{\cdot}$ given contract m_0 may not be unique. (If $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ is not interim efficient, any incentive compatible allocation that Pareto dominates it and gives the agent at least the expected utility of $\mu_0^{\cdot}$ may arise in equilibrium.) We are thus led to define two notions of renegotiation-proofness:

DEFINITION: An allocation $\mu_0^{\cdot}$ corresponding to DRM m_0 is *weakly renegotiation-proof* if there exists an equilibrium of the renegotiation game in which all types of principal propose contract m_0.

[24] See Dewatripont (1986), Hart-Tirole (1988), Laffont-Tirole (1990), and Dewatripont-Maskin (1989) for renegotiation with adverse selection; Fudenberg-Tirole (1990) for moral hazard; and Maskin-Moore (1987) and Aghion-Dewatripont-Rey (1989) for the case of symmetric information.

[25] The concept of a "renegotiation-proof contract" (see below for formal definitions) differs from the "durable decision rule" of Holmström and Myerson (1983). Roughly, a durable decision rule is a contract with the property that any exogenously proposed alternative mechanism would be vetoed (where one veto suffices to stick to the initial decision rule). The exogeneity implies that, in general, only outsiders can lead the renegotiation. Furthermore, Holmström and Myerson assume that the parties do not update their beliefs in case of veto. This latter assumption is problematic because the parties' beliefs, in general, affect the way they will execute the original contract and their payoffs from that contract. These payoffs, in turn, influence the decision whether or not to veto.

[26] Notice that in our framework we allow only one opportunity for renegotiation. Thus, if the agent rejects the principal's renegotiating proposal, she cannot make another offer.

DEFINITION: An allocation $\mu_0^{\cdot}$ corresponding to DRM m_0 is *strongly renegotiation-proof* if it is weakly renegotiation proof and there exists no equilibrium of the renegotiation game in which the contract is renegotiated (i.e., in which the equilibrium outcome is other than $\mu_0^{\cdot}$).

Thus, a strongly renegotiation-proof contract is certain not to be renegotiated, whereas one that is only weakly renegotiation-proof may or may not be, depending on which equilibrium of the renegotiation game is selected. The next two propositions demonstrate that the sets of weakly and strongly renegotiation-proof allocations have a simple structure.

PROPOSITION 8: *An allocation is weakly renegotiation-proof if and only if it is weakly interim efficient.*

PROOF: From the corollary to Proposition 1, a weakly interim-efficient allocation $\mu_0^{\cdot}$ is an RSW allocation relative to itself. Theorem 1 implies that for initial allocation $\mu_0^{\cdot}$, there exists an equilibrium of the renegotiation game in which the principal does not renegotiate, i.e., each type proposes the DRM $\mu_0^{\cdot}$. Conversely, Proposition 5 implies that an allocation that is not weakly interim-efficient is necessarily renegotiated. *Q.E.D.*

PROPOSITION 9: *An allocation $\mu_0^{\cdot}$ is strongly renegotiation-proof only if it is interim-efficient relative to the prior beliefs $\Pi^{\cdot}$. Moreover the converse also holds provided that $\mu_0^{\cdot} = \mu^{\cdot}$ if $\mu^{\cdot}$ is an allocation that is interim-efficient for $\Pi^{\cdot}$ and for which $V^i(\mu^i) = V^i(\mu_0^i)$ for all i.*

PROOF: Consider first an allocation $\mu_0^{\cdot}$ that is not interim-efficient relative to $\Pi^{\cdot}$. Then there exists an incentive compatible allocation that Pareto dominates it and is individually rational for beliefs $\Pi^{\cdot}$. From Theorem 1, this allocation is an equilibrium outcome of the renegotiation game. Hence, $\mu_0^{\cdot}$ is not strongly renegotiation-proof.

Next, let $\mu_0^{\cdot}$ be an initial allocation that is interim-efficient for beliefs $\Pi^{\cdot}$. From Theorem 1, any equilibrium allocation $\mu^{\cdot}$ for the renegotiation game weakly Pareto dominates $\mu_0^{\cdot}$. But because $\mu_0^{\cdot}$ is interim efficient for $\Pi^{\cdot}$, we thus have $V^i(\mu^i) = V^i(\mu_0^i)$ for all i, and so, by hypothesis $\mu^{\cdot} = \mu_0^{\cdot}$.[27] Hence, $\mu_0^{\cdot}$ is strongly renegotiation-proof. *Q.E.D.*

In accordance with Sections 2 through 4, we have concentrated so far on renegotiation where the informed party proposes the new contract (i.e., she leads the renegotiation). But we equally can examine the same game with the roles reversed (i.e., the uninformed party leads the renegotiation). A straightforward, but important corollary of Proposition 9 is as follows.

[27] This reasoning is reminiscent of the no-trade result in Milgrom-Stokey (1982) for initial allocations that are interim-efficient.

PROPOSITION 10: *The strong and weak renegotiation concepts coincide when the uninformed party leads the renegotiation. Under the hypothesis of Proposition 9, an allocation is strongly renegotiation-proof when the informed party leads the renegotiation if and only if it is renegotiation-proof when the uninformed party leads the renegotiation.*

That is, if strong renegotiation-proofness is the "appropriate" version of renegotiation-proofness, it does not matter whether the informed or uninformed party has the bargaining power at the renegotiation stage. The set of ex-ante implementable allocations is the same in both cases.

PROOF OF PROPOSITION 10: When the uninformed party proposes the contract, he chooses an allocation that maximizes his expected payoff for beliefs $\Pi^{\cdot}$ $(\sum_i \Pi^i U^i(\mu^i))$ given the informed party's individual rationality constraints $(V^i(\mu^i) \geqslant V^i(\mu_0^i)$ for all $i)$ and incentive compatibility constraints $(V^i(\mu^i) \geqslant V^i(\mu^j)$ for all i and j). His choice is thus interim-efficient. Conversely, from the reasoning of Proposition 9, an interim-efficient allocation is (strongly) renegotiation-proof regardless of the bargaining process. *Q.E.D.*

We have focused so far on initial allocations that result from direct revelation mechanisms. Although such contracts are especially appealing (because of their simplicity), we can readily consider arbitrary initial mechanisms.

Suppose that we redefine an allocation $\mu_0^{\cdot}$ to be *strongly renegotiation-proof* (SRP) if there exists a contract m_0 whose unique equilibrium outcome is $\mu_0^{\cdot}$ and such that, if m_0 is the initial contract, m_0 is not renegotiated in any equilibrium. Clearly, any allocation $\mu_0^{\cdot}$ that was SRP under the earlier definition (i.e., any interim-efficient allocation) remains so, since we can always take m_0 to be the direct revelation mechanism $\mu_0^{\cdot}$.[28] Moreover, it is easy to see that the new definition does not admit any new SRP allocations. Indeed, suppose $\mu_0^{\cdot}$ is an SRP allocation and let m_0 be an initial contract relative to which $\mu_0^{\cdot}$ is the unique equilibrium outcome of the renegotiation game. If contrary to the claim, $\mu_0^{\cdot}$ is not interim-efficient, then there exists a Pareto-dominating and incentive-compatible allocation $\mu^{\cdot}$ that is individually rational for prior beliefs $\Pi^{\cdot}$. We shall construct an equilibrium of the renegotiation game in which the equilibrium outcome is $\mu^{\cdot}$, a contradiction. Specifically, on the equilibrium path, let all types of principal propose the direct revelation mechanism associated with $\mu^{\cdot}$ and let the agent accept this proposal. If the principal proposes any other contract (including m_0), assign the corresponding continuation equilibrium from the equilibrium giving rise to $\mu_0^{\cdot}$. Such a deviation is, therefore, deterred since by deviating the type i principal gets at most $V^i(\mu_0^i)$ $(\leqslant V^i(\mu^i))$. Summarizing, we have the following proposition.

[28] This is not to say, however, that just because a contract gives rise to an interim-efficient allocation, it will not be renegotiated. We are making the assertion only for direct revelation mechanisms.

PROPOSITION 9*: *An allocation is SRP in the redefined sense if and only if it is SRP in the original sense.*

The story is quite different when we consider general initial contracts and weak renegotiation-proofness. Redefine an allocation $\mu_0^{\cdot}$ to be *weakly renegotiation-proof* (WRP) if there exists a contract m_0 and an allocation $\mu_0^{\cdot}$ and such that, if m_0 is the initial contract, $\mu_0^{\cdot}$ is an equilibrium outcome of the renegotiation game. As with strong renegotiation-proofness, any allocation that was WRP in the original sense (i.e., a weakly interim efficient allocation) is WRP with respect to general initial contracts. However, an allocation need not be WIE to be WRP in the new sense. This is because a general initial contract could have equilibrium outcomes other than $\mu_0^{\cdot}$ that can be used as "threats" to prevent even a highly inefficient allocation from being renegotiated. Indeed, if there exists a desirable good that can be transferred from one party to another in unlimited quantities (as when the Sorting Assumption is imposed), weak renegotiation-proofness is unrestrictive:

PROPOSITION 8*: *Under parts* (*i*) *and* (*ii*) *of the Sorting Assumption, any incentive compatible allocation* $\mu_0^{\cdot}$ *is weakly renegotiation-proof in the redefined sense.*

PROOF: See Appendix C.

We should stress that, although Proposition 8* shows that considerable inefficiency may occur despite the possibility of renegotiation, such inefficiency depends crucially on the choice of a particular continuation equilibrium in the renegotiation game should the principal propose an alternative contract. Thus, although inefficiency *may* arise, it is by no means guaranteed.

B. *Refinements*

As in Section 5, we apply the refinement of Cho-Kreps (1987) (CK) and that of Farrell (1985) and Grossman-Perry (1986) (FGP). An incentive compatible allocation $\mu^{\cdot}$ is "weakly Cho-Kreps renegotiation proof" (weakly CK-RP) if there exists an equilibrium of the renegotiation game (in which the principal leads the renegotiation) that passes the Cho-Kreps intuitive criterion and results in $\mu^{\cdot}$. It is "strongly Cho-Kreps renegotiation proof" (strongly CK-RP) if there exists a unique equilibrium of the renegotiation game that passes the Cho-Kreps intuitive criterion, and if this equilibrium results in $\mu^{\cdot}$. The definitions of "weakly-" and "strongly Farrell-Grossman-Perry renegotiation proof" allocations (weakly- and strongly FGP-RP allocations) are analogous.

Refinements of PBE cannot expand the set of weakly renegotiation proof allocations, because they make it harder to sustain an equilibrium; nor, as Proposition 11 establishes below, does the CK or FGP refinement reduce the

TABLE I[a]

Renegotiation Concept	Equilibrium Concept PBE	CK	FGP
Weakly RP	WIE	WIE	IE[c]
Strongly RP	IE	WIE[b]	IE[c]

[a] RP = renegotiation proof, PBE = perfect Bayesian equilibrium; CK = Cho-Kreps equilibrium; FGP = Farrell-Grossman-Perry equilibrium; WIE = weakly interim efficient; IE = interim efficient relative to beliefs $\Pi^{\cdot}$.
[b] If either (i) $n = 2$ and we ignore all allocations on the boundary of the feasible set or both (ii) the Sorting Assumption holds and (iii) reservation utilities are nondecreasing.
[c] If either (i) or (ii) holds.

set of strongly renegotiation-proof allocations under the hypotheses of Proposition 7. Hence, the pertinent question is whether a refinement induces a greater coincidence of the sets of weakly and strongly renegotiation proof allocations than does unrefined PBE. As Proposition 11 shows, in fact, exact coincidence obtains for the CK and FGP refinements.

PROPOSITION 11: *Suppose that for any WIE allocation $\mu^{\cdot}$ there exist strictly positive beliefs relative to which $\mu^{\cdot}$ is IE. The sets of weakly and strongly renegotiation-proof allocations* (*when the initial contract is a DRM*) *for unrefined PBE and the Cho-Kreps and FGP refinements are given in Table I.*

PROOF: See Maskin-Tirole (1990b).

Proposition 11 sheds light on a result due to Nosal (1988). Nosal considers a model in which two parties sign an enforceable contract under symmetric information. One of the two parties then receives private information and can offer to renegotiate the contract. Nosal shows that the optimal contract when renegotiation is prohibited is robust when renegotiation is feasible. The link between this result and Proposition 11 is that an optimal contract in the absence of renegotiation is necessarily interim-efficient and therefore (strongly) renegotiation proof from Proposition 11.

The reason why the CK criterion has no power to reduce the set of weakly renegotiation proof allocations is easily grasped from Figure 4, which depicts the compensation example with $n = 2$ and $U_0^i = \bar{U} = 0$ for all i. The allocation $\{\mu_*^1, \mu_*^2\}$ is weakly interim efficient (μ_*^1 is *ex-post* efficient, and therefore cannot be improved on for the type-1 principal without lowering U^1; and μ_*^2 cannot be improved upon for the type-2 principal without violating incentive compatibility or lowering U^2, as the shaded region is to the northwest of the line $U^2(\mu^2) = U^2(\mu_*^2)$). Note also that $U^2(\mu_*^2) > 0$ and $U^1(\mu_*^1) < 0$. Suppose that $\{\mu_*^1, \mu_*^2\}$ is a candidate equilibrium allocation, but that the principal proposes the outcome μ^2 (see Figure 4) instead. Note that $V^2(\mu^2) > V^2(\mu_*^2)$

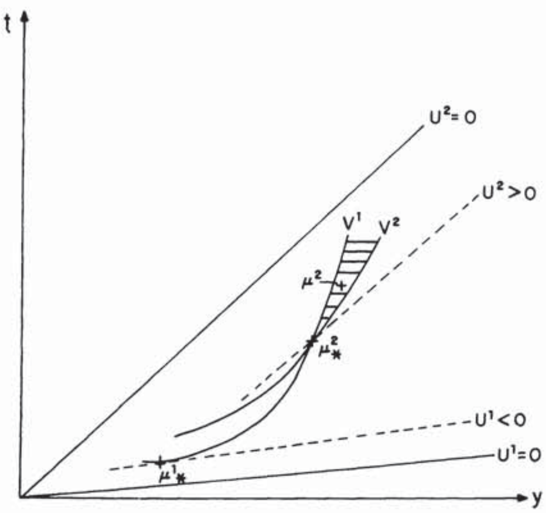

FIGURE 4

and $V^1(\mu^2) < V^1(\mu^1_*)$. Hence, the CK criterion postulates that the agent when confronting contract proposal μ^2 should believe that the principal's type is 2. If the parties are not yet bound by an initial contract (as in Section 5), the agent accepts proposal μ^2 since $U^2(\mu^2) > 0$, and thus the candidate equilibrium $\{\mu^1_*, \mu^2_*\}$ is upset. Suppose, by contrast, that the two parties have already signed contract $\{\mu^1_*, \mu^2_*\}$. The no-renegotiation equilibrium passes the CK criterion: Proposing μ^2 "convinces" the agent that the principal has type 2, but this new contract is rejected by the agent since $U^2(\mu^2) < U^2(\mu^2_*)$.

Let us briefly consider how Proposition 11 is affected when the initial contract can be a general mechanism rather than only a DRM. From Proposition 8* the set of weakly renegotiation-proof allocations is just the set of incentive compatible allocations. The other entries in Table I are unchanged, except for the set of the CK-WRP allocations, which, we conjecture, coincides with the set of incentive compatible allocations.

7. SIGNALING VS. SCREENING

Most of the literature on markets with adverse selection (since Rothschild-Stiglitz (1976) and Wilson (1977)) assumes that uninformed parties propose contracts. Let us refer to this as the *screening* approach. Often (in analyses of competitive markets in particular) there are at least two uninformed parties (UP)—e.g., employers, insurance companies ... —who compete with each other in Bertrand fashion. The informed party (IP)—employee, insurance customer ... —chooses the contract for which her utility is highest.

In contrast to screening, Sections 2 through 5 of this paper analyze a *signaling* model in which the informed party proposes the contract (note that all the results of Sections 2 through 5 go through with more than one uninformed party). Whether or not the set of equilibria depends on who proposes the contract is the focus of this section.[29]

To study equilibrium with screening, we assume that there are (at least) two UP's (but only one IP). An UP obtains utility $U^i(\mu)$ if he signs a contract (resulting in outcome μ) with the IP (and the latter's type is type i). His utility is $\bar{U}$ if he fails to sign a contract. The UP's simultaneously propose contracts to the IP. A contract is, as before, simply a game form m between the two parties that results in an outcome. The IP (of type i) obtains utility $V^i(\mu)$ if she accepts a contract that results in outcome μ; she gets her reservation utility $\bar{V}^i$ if she rejects all contracts. The uninformed parties have prior beliefs $\Pi^{\cdot}$ about the IP's type.[30]

We first show that the equilibrium set of this screening game is very large, but then note that some equilibria are not robust to a simple modification of the model, where the informed party is given some influence over the contract. We conclude that the equilibrium allocations in this modified model coincide with those in the signaling model of Section 5.

PROPOSITION 12: *Suppose that condition (iv) of Proposition 7 holds. Any allocation $\mu^{\cdot}=\{\mu^i\}_{i=1}^n$ that satisfies the informed party's incentive constraints ($V^i(\mu^i) \geqslant V^i(\mu^j)$ for all i and j), satisfies her individual rationality constraints ($V^i(\mu^i) \geqslant \bar{V}^i$ for all i), and breaks even for the uninformed parties ($\sum_i \Pi^i U^i(\mu^i) = \bar{U}$ for prior beliefs $\Pi^{\cdot}$) is an equilibrium outcome of the screening model.*

PROOF: Let A be the set of allocations (including the null allocation) that satisfy the conditions of Proposition 12. Consider the contract m^*, which specifies that first the UP is free to choose any allocation $\mu^{\cdot} \in A$ and then the IP gets to choose from the menu $\{\mu^i\}$.[31] We claim that for any $\bar{\mu}^{\cdot} \in A$, there exists an equilibrium in which all UP's propose m^*; the IP accepts one of the offers; the UP concerned then chooses $\bar{\mu}^{\cdot}$; and, finally, the type i IP chooses $\bar{\mu}^i$. In this equilibrium an uninformed party's beliefs if his proposal is accepted are the priors $\Pi^{\cdot}$, as long as one of the UP's has proposed m^*.

Because $\bar{\mu}^{\cdot}$ is incentive compatible it is indeed optimal for the type i IP to choose $\bar{\mu}^i$, and, because the UP is indifferent among all allocations in A (given beliefs $\Pi^{\cdot}$) he might as well choose $\bar{\mu}^{\cdot}$. Moreover, the individual rationality of $\bar{\mu}^{\cdot}$ ensures that the IP is willing to accept m^*.

It remains to construct out-of-equilibrium behavior so that no UP gains from proposing $m \neq m^*$. Let $\mu^{\cdot}$ denote an equilibrium allocation (for beliefs $\Pi^{\cdot}$) of the game form in which first the IP chooses between m and the null allocation

[29] For a comparison of screening and signaling models when the screening variable (y) is chosen before contracting, see Madrigal-Tan (1986) and Stiglitz-Weiss (1983).

[30] Hellwig (1986) considers a similar game, except that he constrains contracts to belong to a particular class: A contract consists, first, of an allocation (i.e., direct revelation mechanisms). After the IP accepts the contract and chooses from the menu, the UP who proposed it has the right to revert to the reservation allocation, i.e., to withdraw his offer.

[31] This contract does not belong to the class M of Section 2.B because it is neither finite nor a simultaneous-move game. However, it satisfies the equilibrium existence and upper hemicontinuity properties that led us to reduce to M in the first place.

and then, if she chooses the former, the two parties play m. There are two cases. If $\sum_i \Pi^i U^i(\mu^i) \leqslant \bar{U}$, then we can construct a continuation equilibrium, starting from the point where one UP has deviated by proposing m, in which all types of IP either accept m or reject all contracts, resulting in the allocation $\mu^{\cdot}$ for the deviator. If the IP chooses one of the other UP's proposals, m^*, then that UP maintains his prior beliefs $\Pi^{\cdot}$ and chooses the null allocation. Clearly the UP who proposes m does not gain from the deviation. If $\sum_i \Pi^i U^i(\mu^i) > \bar{U}$, then suppose that if some UP proposes m, the others (who have proposed m^*) choose an allocation $\tilde{\mu}^{\cdot}$ in A that strictly Pareto dominates $\mu^{\cdot}$ (such an allocation exists from condition (iv) of Proposition 7, because the uninformed party's individual rationality constraint for $\mu^{\cdot}$ is not binding) if their proposal is accepted. Then for all i, $V^i(\tilde{\mu}^i) > V^i(\mu^i)$, which means that there exists a continuation equilibrium in which all types of the IP choose m^*, and so the deviation again is not profitable. *Q.E.D.*

Proposition 12 shows that the equilibrium set of this screening model includes allocations for which the IP's payoff is actually lower than in the RSW allocation for the null trade. The possibility of such low payoffs, however, is a knife-edge result that depends on the IP's having no power to influence the contract proposal. Intuitively, if she had even only slight influence, she ought to be able to exploit the Bertrand competition between the UP's to attain her RSW payoff.

One way of formalizing this intuition is to suppose that there is a large number N of uninformed parties who play with the informed party the following variant of the Rubinstein (1982) bargaining game. At each date $t = 1, \ldots, m_I$, the informed party makes a contract proposal to the uninformed parties (and chooses randomly among them if several accept). Then at dates $t = m_{I+1}, \ldots, m_I + m_U$ the uninformed parties make simultaneous contract proposals to the informed party (who accepts one of them or rejects them all). Then at dates $m_I + m_U + 1 \ldots, 2m_I + m_U$, the informed party makes proposals again, etc. The parties discount the future with discount factor $\delta = e^{-rT}$ per period (where T is the interval between bargaining dates). The game ends once a contract proposal is accepted. The payoffs to the informed party and the chosen uninformed party are $\delta^t V^i(\mu)$ and $\delta^t U^i(\mu)$, where t is the date of agreement and μ the outcome resulting from the contract. Suppose that bargaining occurs quickly ($T \to 0$). Note that if m_U/m_I is large, the informed party has perhaps little bargaining power.[32]

We claim that in this bargaining game the informed party can closely approximate her RSW payoff. Suppose that, when it is her turn, she proposes an incentive-compatible allocation near the RSW allocation but where the UP's payoff is $\bar{U} + \varepsilon$ regardless of the IP's type. An uninformed party's utility from

[32] In the two-player, symmetric information version of Rubinstein (1982), the more proposals a party can make, the bigger his share of the pie. Thus by assuming m_U/m_I large, we endow the informed party with only "slight power" to influence the contract.

accepting this proposal is $\bar{U}+\varepsilon$. However, there is an upper bound on how much utility the uninformed parties can obtain in the aggregate if everyone rejects this proposal (because of the informed party's individual rationality constraint). Therefore in any continuation equilibrium corresponding to rejection, at least one UP must obtain less than $\bar{U}+\varepsilon$ because he has many competitors ("many competitors" is a large number that depends on ε, but not on T). Thus he will accept the informed party's proposal, and so the RSW payoff is an approximate lower bound for the informed party when she proposes a contract. As long as T is near 0, this payoff is also an approximate lower bound at the beginning of the bargaining game.

Conversely, one can show that any incentive-compatible allocation that is individually rational (for beliefs $\Pi^{\cdot}$) and Pareto dominates the RSW allocation for the null trade allocation is an equilibrium of the bargaining game. The proof is essentially a combination of those for Theorem 1 and Proposition 12. We thus conclude with Proposition 13.

PROPOSITION 13: *For any $\varepsilon > 0$, there exist T sufficiently small and a number of uninformed parties sufficiently big so that if $\mu^{\cdot}$ is an equilibrium of the bargaining game, $V^i(\mu^i) \geqslant \hat{V}^i(\mu_0^i) - \varepsilon$ for all i. Conversely, for any T, any allocation characterized in Proposition 6 is an equilibrium allocation of the bargaining game.*

8. BILATERAL ASYMMETRIC INFORMATION

Up to now we have assumed that the agent—the party who accepts or rejects the contract proposal—has no private information himself. However, the case of bilateral asymmetric information is of course important in practice. A firm with private information about its environment or technology may offer a labor contract to an employee with private information about his ability, or the reverse. A manufacturer with private information about her product's quality may price discriminate among consumers with different preferences. Or a franchiser with information about quality and therefore aggregate demand for her product may offer a contract to a franchisee with private information about his talent or about local demand. There is an important special case in which many of our results continue to hold even when the agent has private information. The leading feature of this case is that, in addition to the Sorting Assumption being satisfied, parties have *quasi-linear* utility functions.

This section is organized as follows. We first generalize the notions of weakly interim efficient (WIE*), Rothschild-Stiglitz-Wilson (RSW*), and interim efficient (IE*) allocations to bilateral asymmetric information. We then note that Theorem 1, the assertion that the equilibrium allocations are exactly those that dominate the RSW allocation and are incentive compatible and individually rational (for prior beliefs $\Pi^{\cdot}$) provided that the RSW allocation is IE for some strictly positive beliefs $\hat{\Pi}^{\cdot}$, carries over to bilateral asymmetric information. The problem is then to find sufficient conditions that guarantee that the RSW* allocation is IE* for some such beliefs. At this point, we specialize the model to quasi-linear utility functions for the principal and the agent, and show that if the agent's binding incentive compatibility constraints in the RSW* program are the "downward adjacent constraints," then the RSW* allocation is indeed IE* for some strictly positive beliefs. The last step consists of invoking the Sorting Assumption to ensure that the agent's binding constraints are the downward incentive compatibility constraints.

A. *Efficiency Concepts*

Let there be n types of principal $(i = 1, \ldots, n)$ and m types of agent $(j = 1, \ldots, m)$. The probabilities of type i of principal and type j of agent are Π^i and p_j, respectively. The principal's and the agent's types are independently distributed. An allocation in our expanded framework is a menu $\mu_{\cdot}^{\cdot} = \{\mu_j^i\}_{i,j}$ of outcomes, one for each pair of principal's and agent's types. The utility functions are also indexed by the two types: $V_j^i(\mu)$ for the principal and $U_j^i(\mu)$ for the agent, for a given outcome μ. A reservation or status-quo allocation is denoted $\bar{\mu}_{\cdot}^{\cdot}$. An allocation $\mu_{\cdot}^{\cdot}$ is incentive compatible for the principal, if for all i,

$$(\mathrm{PIC}^i) \quad \sum_j p_j V_j^i(\mu_j^i) \geqslant \sum_j p_j V_j^i(\mu_j^k) \qquad \text{for all } k.$$

For the agent, we define incentive compatibility and individual rationality both "type-by-type" (i.e., for each type of principal) and "on average" (i.e., in expectation over the principal's type for some beliefs). The type-by-type concepts are (for all i)

$$(\mathrm{AIC}_j^i) \quad U_j^i(\mu_j^i) \geqslant U_j^i(\mu_k^i) \qquad \text{for all } j \text{ and } k,$$

$$(\mathrm{IR}_j^i(\bar{\mu}_{\cdot}^{\cdot})) \quad U_j^i(\mu_j^i) \geqslant U_j^i(\bar{\mu}_j^i) \qquad \text{for all } j.$$

For arbitrary beliefs $\hat{\Pi}^{\cdot}$ about the principal's type, the on-average concepts are

$$(\mathrm{AIC}_j(\hat{\Pi}^{\cdot})) \quad \sum_i \hat{\Pi}^i U_j^i(\mu_j^i) \geqslant \sum_i \hat{\Pi}^i U_j^i(\mu_k^i) \qquad \text{for all } j \text{ and } k,$$

$$(\mathrm{IR}_j(\hat{\Pi}^{\cdot}, \bar{\mu}_{\cdot}^{\cdot})) \quad \sum_i \hat{\Pi}^i U_j^i(\mu_j^i) \geqslant \sum_i \hat{\Pi}^i U_j^i(\bar{\mu}_j^i) \qquad \text{for all } j.$$

Clearly, (AIC_j^i) for all i implies $(\mathrm{AIC}_j(\hat{\Pi}^{\cdot}))$ for any $\hat{\Pi}^{\cdot}$, and $(\mathrm{IR}_j^i(\bar{\mu}_{\cdot}^{\cdot}))$ for all i implies $(\mathrm{IR}_j(\hat{\Pi}^{\cdot}, \bar{\mu}_{\cdot}^{\cdot}))$ for any $\hat{\Pi}^{\cdot}$.

Let us confine attention to initial allocations $\mu_{\cdot 0}^{\cdot}$ that satisfy (PIC^i) and (AIC_j^i) for all i and j. As in the one-sided private information model, $\mu_{\cdot 0}^{\cdot}$ can be a "no-trade" allocation, or else might result from a previous, incentive compatible contract.

DEFINITION: An allocation $\bar{\mu}^{\cdot}_{\cdot}$ is WIE* if and only if it is a solution to Program I* for some vector of positive weights $\{w^i\}_{i=1,\ldots,n}$:

$$\text{Program I}^*: \max_{\mu^{\cdot}_{\cdot}} \sum_i w^i \sum_j p_j V_j^i\left(\mu_j^i\right)$$

subject to (PIC^i) for all i, (AIC_j^i) for all i and j, and ($\text{IR}_j^i(\bar{\mu}^{\cdot}_{\cdot})$) for all i and j.

As before, the agent's constraints in the definition of weak interim efficiency are imposed type by type. The same is true of the (generalized) concept of Rothschild-Stiglitz-Wilson allocation:

DEFINITION: An allocation $\hat{\mu}^{\cdot}_{\cdot}(\mu^{\cdot}_{\cdot 0})$ is RSW* relative to $\mu^{\cdot}_{\cdot 0}$ if and only if for all i, $\hat{\mu}^i_{\cdot}(\mu^{\cdot}_{\cdot 0}) = \mu^i_{\cdot}$, where $\mu^{\cdot}_{\cdot}$ solves Program II^{i*}:

$$\text{Program II}^{i*}: \max_{\mu^{\cdot}_{\cdot}} \sum_j p_j V_j^i\left(\mu_j^i\right)$$

subject to (PIC^k) for all k, (AIC_j^k) for all k and j, and ($\text{IR}_j^k(\mu^{\cdot}_{\cdot 0})$) for all k and j.

Once again, the definition is the same as in the case where the agent has no private information except for the presence of the agent's type-by-type incentive compatibility constraints. The assumption that $\mu^{\cdot}_{\cdot 0}$ satisfies the principal's and the agent's type-by-type incentive compatibility constraints ensures that the constraints in II^{i*} can all be satisfied.

From the same argument as in the proof of Proposition 1, $\hat{\mu}^{\cdot}_{\cdot}(\mu^{\cdot}_{\cdot 0})$ is WIE* and, in particular, incentive compatible for the principal. It therefore solves Program II* for any set of positive weights $\{w^i\}_{i=1,\ldots,n}$:

$$\text{Program II}^*: \max_{\mu^{\cdot}_{\cdot}} \sum_i w^i \left(\sum_j p_j V_j^i\left(\mu_j^i\right)\right)$$

subject to (PIC^i) for all i, (AIC_j^i) for all i and j, and ($\text{IR}_j^i(\mu^{\cdot}_{\cdot 0})$) for all i and j.

DEFINITION: An allocation $\bar{\mu}^{\cdot}_{\cdot}$ is IE* relative to beliefs $\hat{\Pi}^{\cdot}$ if and only if for some vector of positive weights $\{w^i\}_{i=1,\ldots,n}$, it solves program

$$\text{Program V}^*: \max_{\mu^{\cdot}_{\cdot}} \sum_i w^i \left(\sum_j p_j V_j^i\left(\mu_j^i\right)\right)$$

subject to (PIC^i) for all i, ($\text{AIC}_j(\hat{\Pi}^{\cdot})$) for all j and ($\text{IR}_j(\hat{\Pi}^{\cdot}, \bar{\mu}^{\cdot}_{\cdot})$) for all j.

Note that in the definition of IE* allocations, the agent's constraints hold in expectation. Although the definitions of WIE* and IE* allocations are natural generalizations of WIE and IE allocations, the reader should note two points. First, the notion of IE* allocations is not quite the usual notion of interim efficiency, which would require that all types of agent have positive weight in the objective function. (This distinction does not arise when the agent has no private information because the IR constraint in the definition of IE could be reformulated by including the agent in the objective function with a positive weight.) Second, unlike the case where the agent has no private information, an allocation that is IE* need not be WIE*. This is because the constraints $(\text{AIC}_j(\hat{\Pi}^\cdot))$ in Program V* do not ensure that the constraints (AIC_j^i) in Program I* are satisfied. It is precisely this failure that, in general, prevents us from straightforwardly extending our previous results to the case of an agent with private information. Intuitively, when the incentive constraints need hold only on average (with respect to beliefs $\Pi^\cdot$)—as in Program V*—the principal may gain from violating some of the *individual* constraints (AIC_j^i), which she can do so long as the violations are made up by tightening other (AIC_j^i) constraints so that $(\text{AIC}_j(\Pi^\cdot))$ holds. These trade-offs between relaxing and tightening constraints are the focus of Maskin-Tirole (1990a). In that companion piece, however, we show that when the utility functions are quasi-linear the principal derives no benefit from violating any of the (AIC_j^i) constraints. That suggests, and below we confirm, that much of the analysis in Sections 2 through 7 extends with quasi-linearity.

B. *Equilibrium Outcomes*

THEOREM 1*: *Suppose that the RSW* allocation $\hat{\mu}^\cdot_\cdot(\mu^\cdot_{\cdot 0})$ is IE* for some strictly positive beliefs $\hat{\Pi}^\cdot$ (i.e., $\hat{\Pi}^i > 0$ for all i). Then the set of equilibrium allocations of the contract proposal game is the set of allocations $\mu^\cdot_\cdot$ that satisfy (PIC^i) for all i, $(AIC_j(\Pi^\cdot))$ for all j, and $(IR_j(\Pi^\cdot))$ for all j, and that weakly Pareto dominate the RSW* allocation: for all i,*

$$\sum_j p_j V_j^i\left(\mu_j^i\right) \geqslant \sum_j p_j V_j^i\left(\hat{\mu}_j^i(\mu^\cdot_{\cdot 0})\right).$$

The proof of Theorem 1* is identical to that of Theorem 1. The only formal change is that the last step of the proof is performed for the agent's incentive compatibility as well as individual rationality constraints.

REMARK: When the agent has no private information, an RSW allocation is necessarily IE with respect to some beliefs (although more assumptions are needed to make sure that these beliefs can be chosen strictly positive). But with bilateral asymmetry, an RSW* allocation need not be IE* with respect to any beliefs.

C. *The Quasi-Linear Case*

We now make an assumption that ensures that an RSW* allocation is IE* for some strictly positive beliefs and, consequently, that Theorem 1* characterizes the equilibrium outcomes of the contract proposal game.

ASSUMPTION Q: (i) *Preferences are quasi-linear: for all i and j,*

$$V_j^i(y,t) = t - \phi_j^i(y),$$
$$U_j^i(y,t) = \psi_j^i(y) - t.$$

(ii) *In Program II*, the binding constraints among the agent's constraints* (AIC_j^i) *and* $(IR_j^i(\mu_{\cdot 0}^{\cdot}))$ *are the downward adjacent incentive constraints and the type-*1 *individual rationality constraint. That is, for all i:*

$$U_j^i(\mu_j^i) \geqslant U_j^i(\mu_{j-1}^i) \qquad (j = 2, \dots, m)$$

and

$$U_1^i(\mu_1^i) \geqslant U_1^i(\mu_{10}^i).$$

Furthermore, the Lagrange multipliers associated with these constraints are strictly positive.

PROPOSITION 14: *Under Assumption Q,*[33] *the RSW* allocation* $\hat{\mu}^{\cdot}_{\cdot}(\mu^{\cdot}_{\cdot 0})$ *is IE* for some strictly positive beliefs* $\hat{\Pi}^{\cdot}$ *(and therefore Theorem* 1* *applies).*

PROOF: See Maskin-Tirole (1990b).

To sum up, from Theorem 1*, the results of this paper carry over to bilateral asymmetric information if the RSW* allocation is IE* for some strictly positive beliefs. Under quasi-linear preferences, this hypothesis is satisfied if the set of binding constraints consists of the agent's downward adjacent IC constraints and IR constraint at the bottom. Here, we have not attempted to provide the most general conditions under which these are the binding constraints (although there are many examples from the literature where they indeed are). Rather, we focus on the case where each party has two possible types and the Sorting Condition is satisfied (see Appendix D). We conjecture that the analysis carries over to much more general environments satisfying the Sorting Condition.

9. *Comparison with Myerson* (1983)

Let us compare our results to those of Myerson (1983). Myerson analyzes contract design by an informed party (player 1—the principal) in a more general context than ours. In particular, the other parties (players 2 through N

[33] Whether or not Assumption Q is satisfied can be easily checked in particular applications. A good strategy for doing so may consist of:

(i) *presuming* that in Program II*, the binding constraints are the principal's *upward* adjacent incentive constraints (type i announcing type $(i+1)$), the agent's *downward* incentive constraints (type j announcing type $(j-1)$ for each i), and individual rationality constraint at the bottom;

(ii) seeing whether y_j^i is monotonic in i and j and the solution satisfies the omitted constraints. We conjecture, but have not proved, that assumptions (b) through (e) in Proposition 2* in Appendix D together with a monotone hazard rate condition on the distribution of the agent's type imply Assumption Q. We will content ourselves with showing that assumptions (b) through (e) imply Assumption Q when there are two types of principal and agent and when the agent's information does not enter the principal's utility function.

—the agents) may also have private information even when utility is not quasi-linear. Myerson defines an allocation to be "incentive compatible type-by-type" or "safe" if it is incentive compatible when the agents know the principal's true type. A "strong solution" is an allocation that is safe and interim-efficient (i.e., undominated for the different types of principal in the class of incentive compatible allocations). In our restricted framework, incentive compatibility and incentive compatibility type-by-type amount to the agent's being willing to accept to play the mechanism (individual rationality) when he has prior beliefs about, and he knows, the principal's type, respectively. Thus for an initial allocation $\mu_0^{\cdot}$, the RSW allocation $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ is safe by definition, and is a strong solution if and only if it is interim-efficient.[34]

Myerson then briefly analyzes the noncooperative game in which the principal offers a mechanism (Section 5) and shows that an equilibrium exists and that any strong solution, if one exists, is an equilibrium. These results mean in our context that $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ is an equilibrium outcome if it is interim-efficient. Myerson's general observation that equilibria are not generally unique even when a strong solution exists (p. 1781) does not apply to our more structured model; our Theorem 1 indicates that $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ is the unique equilibrium outcome when it is interim-efficient, i.e., when a strong solution exists. Our Theorem 1 also fully characterizes the equilibrium set when the RSW allocation is not interim-efficient.

After discussing the noncooperative game, Myerson goes on to develop a cooperative approach. He first identifies a subset of interim-efficient allocations, core allocations. He then isolates a subset of core allocations, neutral allocations, or mechanisms that forms the smallest class satisfying four axioms. In particular, neutral mechanisms must include strong solutions if such exist, and the set of neutral mechanisms cannot expand if the set of public actions (actions that can be contractually specified) becomes larger. He proves existence of and characterizes neutral mechanisms. In our context, when $\hat{\mu}^{\cdot}(\mu_0^{\cdot})$ is interim-efficient, it is the unique neutral solution as it is undominated, and thus is the unique core solution.

Dept. of Economics, Harvard University, Cambridge, MA 02138, U.S.A.
and
Dept. of Economics, MIT, Cambridge, MA 02139, U.S.A.

Manuscript received November, 1988; final revision received January, 1991.

APPENDIX A: Proof of Proposition 2

Let $\{(\hat{y}^k, \hat{t}^k)\}_{k=1}^n$ be a solution to the successive Programs III^k. (Such a solution exists from part (iii) of the Sorting Assumption.) We claim first that $U^k(\hat{y}^k, \hat{t}^k) = U_0^k$ and

(1) $\hat{y}^k < \hat{y}^{k+1}$ for all k.

[34] Interestingly, Myerson's Theorem 1, although not written in a noncooperative set up, is similar to the second part of our Proposition 7, according to which if the RSW is interim-efficient, it is immune to FGP deviations.

The proof is by induction on n. For $n=1$, the claim is vacuously true. Suppose that it is true for $n-1$. It remains to show that it is true for n, i.e., that $\hat{y}^{n-1}<\hat{y}^n$ and $U^n(\hat{y}^n,\hat{t}^n)=U_0^n$. If $\hat{y}^{n-1}=\hat{y}^n$, then optimality and the incentive constraints, and the fact that U^i is increasing in i, imply that $\hat{t}^{n-1}=\hat{t}^n$. Thus, since $U^{n-1}(\hat{y}^{n-1},\hat{t}^{n-1})\geqslant U_0^{n-1}$, we have

$$(2)\qquad U^n(\hat{y}^n,\hat{t}^n)>U_0^n.$$

Inequality (2) implies that we can increase $\hat{y}^n$ and $\hat{t}^n$ slightly to $(\hat{y}^n+\alpha,\hat{t}^n+\beta)$ so that $V^{n-1}(\hat{y}^{n-1},\hat{t}^{n-1})=V^{n-1}(\hat{y}^n+\alpha,\hat{t}^n+\beta)$ and $U^n(\hat{y}^n+\alpha,\hat{t}^n+\beta)>U_0^n$. But from (iv) of the Sorting Assumption, $V^n(\hat{y}^n+\alpha,\hat{t}^n+\beta)>V^n(\hat{y}^n,\hat{t}^n)$, a contradiction of the fact that $(\hat{y}^n,\hat{t}^n)$ solves III^n. Hence $\hat{y}^{n-1}\neq\hat{y}^n$. Because

$$(3)\qquad V^{n-1}(\hat{y}^{n-1},\hat{t}^{n-1})\geqslant V^{n-1}(\hat{y}^n,\hat{t}^n)$$

and V_t^{n-1} is bounded away from zero, we can choose $t\geqslant\hat{t}^n$ such that $V^{n-1}(\hat{y}^{n-1},\hat{t}^{n-1})=V^{n-1}(\hat{y}^n,t)$. From the Sorting Assumption, $V^n(\hat{y}^n,\hat{t}^n)\leqslant V^n(\hat{y}^n,t)<V^n(\hat{y}^{n-1},\hat{t}^{n-1})$ if $\hat{y}^n<\hat{y}^{n-1}$, a contradiction of the fact that $(\hat{y}^n,\hat{t}^n)$ maximizes V^n subject to IC^n and IR$_0^n$ ($(\hat{y}^{n-1},\hat{t}^{n-1})$ is a feasible choice for the type n principal since $U^n(\hat{y}^{n-1},\hat{t}^{n-1})\geqslant U^{n-1}(\hat{y}^{n-1},\hat{t}^{n-1})=U_0^{n-1}\geqslant U_0^n$). Hence $\hat{y}^n>\hat{y}^{n-1}$, as required. This implies immediately from optimality that $\hat{t}^n>\hat{t}^{n-1}$. Above we showed that inequality (2) leads to contradiction. Hence $U^n(\hat{y}^n,\hat{t}^n)=U_0^n$ as well.

We next claim that $\{(\hat{y}^k,\hat{t}^k)\}_{k=1}^n$ is incentive compatible. Again, the proof is by induction. Suppose that the claim has been established for $n-1$. For n, we must show that $V^k(\hat{y}^k,\hat{t}^k)\geqslant V^k(\hat{y}^n,\hat{t}^n)$ and $V^n(\hat{y}^n,\hat{t}^n)\geqslant V^n(\hat{y}^k,\hat{t}^k)$ for all $k\leqslant n-1$. Now, the former inequality follows from part (iv) of the Sorting Assumption and the inequalities (1) and (3). The latter inequality follows from the facts that (i) $(\hat{y}^n,\hat{t}^n)$ maximizes V^n subject to IC^n and IR$_0^n$, (ii) $V^{n-1}(\hat{y}^{n-1},\hat{t}^{n-1})\geqslant V^{n-1}(\hat{y}^k,\hat{t}^k)$ for $k=1,\ldots,n-1$ (by inductive hypothesis), and (iii) $U^n(\hat{y}^{n-1},\hat{t}^{n-1})\geqslant U_0^n$ (since U^i is increasing in i and U_0^i is nonincreasing in i).

Finally, we claim that any solution $\{\bar{y}^k,\bar{t}^k)\}_{k=1}^n$ to

Program IV^n: $\max V^n(y^n,t^n)$

subject to

$$(\text{IC}^k)\qquad V^{k-1}(y^{k-1},t^{k-1})\geqslant V^{k-1}(y^k,t^k)\qquad(k=2,\ldots,n)$$

and

$$(\text{IR}^k)\qquad U^k(y^k,t^k)\geqslant U_0^k\qquad(k=1,\ldots,n)$$

satisfies $V^n(\bar{y}^n,\bar{t}^n)=V^n(\hat{y}^n,\hat{t}^n)$. Once again, we proceed by induction. Assume the claim is true for $n-1$. Consider the solution to Program IV^n for n. By inductive hypothesis $V^{n-1}(\hat{y}^{n-1},\hat{t}^{n-1})$ is the maximized value of $V^{n-1}(\cdot,\cdot)$ subject to the constraints of Program IV^n. Hence

$$(4)\qquad V^{n-1}(\bar{y}^{n-1},\bar{t}^{n-1})\leqslant V^{n-1}(\hat{y}^{n-1},\hat{t}^{n-1}).$$

Now, $(\bar{y}^n,\bar{t}^n)$ solves the following simplified version of Program IV^n:

Program IV$_*^n$: $\max V^n(y^n,t^n)$ subject to

$$(\text{IC}_*^n)\qquad V^{n-1}(\bar{y}^{n-1},\bar{t}^{n-1})\geqslant V^{n-1}(y^n,t^n)$$

and

$$(\text{IR}_*^n)\qquad U^n(y^n,t^n)\geqslant U_0^n.$$

Thus, (4) implies that Program IV$_*^n$ is more constrained than III^n, and so $V^n(\bar{y}^n,\bar{t}^n)\leqslant V^n(\hat{y}^n,\hat{t}^n)$. But $\{(\hat{y}^k,\hat{t}^k)\}$ satisfies all the constraints of Program IV^n, and so $V^n(\bar{y}^n,\bar{t}^n)=V^n(\hat{y}^n,\hat{t}^n)$, as claimed.

Now, Program II^n is more constrained than III^n. But $\{(\hat{y}^k,\hat{t}^k)\}$ solves the latter program, and, because it is incentive compatible, it satisfies all the constraints of the former. Hence, it solves the former program as well. *Q.E.D.*

APPENDIX B: PROOF OF PROPOSITION 4

(a) From Proposition 3 there exist beliefs $\Pi^{\cdot}$ for which $\hat{\mu}^{\cdot}$ is interim efficient. Suppose, contrary to Proposition 4, that $\Pi^i = 0$ for some i. From (iv) of the Sorting Assumption, we can increase $(\hat{y}^i, \hat{t}^i)$ slightly to $(\hat{y}^i + \alpha, \hat{t}^i + \beta)$ so that

$$(5) \qquad V^i(\hat{y}^i + \alpha, \hat{t}^i + \beta) > V^i(\hat{y}^i, \hat{t}^i)$$

but

$$(6) \qquad V^j(\hat{y}^i + \alpha, \hat{t}^i + \beta) < V^j(\hat{y}^j, \hat{t}^j) \qquad \text{for all } j < i.$$

Because $\hat{\mu}^{\cdot}$ is deterministic, the proof of Proposition 2 implies that $(\hat{y}^{i+1}, \hat{t}^{i+1})$ solves

$$\max V^{i+1}(y^{i+1}, t^{i+1}) \qquad \text{subject to}$$

$$V^i(\hat{y}^i, \hat{t}^i) \geqslant V^i(y^{i+1}, t^{i+1})$$

and

$$U^{i+1}(y^{i+1}, t^{i+1}) \geqslant U_0^{i+1}.$$

Now $(y^{i+1}, t^{i+1}) = (\hat{y}^i, \hat{t}^i)$ satisfies the constraints of this program. But, from Proposition 2, $\hat{y}^i < \hat{y}^{i+1}$. Hence from (iv) of the Sorting Assumption, $V^{i+1}(\hat{y}^{i+1}, \hat{t}^{i+1}) > V^{i+1}(\hat{y}^i, \hat{t}^i)$. Similarly,

$$(7) \qquad V^j(\hat{y}^j, \hat{t}^j) > V^j(\hat{y}^i, \hat{t}^i) \qquad \text{for all } j > i.$$

But, in view of (7), we can choose α and β in (5) and (6) so that, in addition, (6) holds for all $j > i$. Then the allocation $\hat{\mu}^{\cdot}$ with $(\hat{y}^i, \hat{t}^i)$ replaced by $(\hat{y}^i + \alpha, \hat{t}^i + \beta)$ is incentive compatible, Pareto dominates $\hat{\mu}^{\cdot}$ and yet generates the same expected utility for the agent (since $\Pi^i = 0$). This violates the interim efficiency of $\hat{\mu}^{\cdot}$ relative to $\Pi^{\cdot}$, and so we conclude that $\Pi^i > 0$ for all i.

(b) We proceed by induction on k. For $k = 1$ the claim is trivially true because, from Proposition 2, $(\hat{y}^1, \hat{t}^1)$ is interim efficient relative to any beliefs. Assume that the claim holds for ${}^{k}\hat{\mu}^{\cdot}$. Suppose that the allocation ${}^{k+1}\hat{\mu}^{\cdot} = \{(\hat{y}^1, \hat{t}^1), \dots, (\hat{y}^{k+1}, \hat{t}^{k+1})\}$ is interim efficient in the $(k+1)$-type model for beliefs $\{\Pi^1, \dots, \Pi^{k+1}\}$ where $\Pi^i > 0$ for all i and $\sum_{i=1}^{k+1} \Pi^i = 1$. If, contrary to the claim, $\{(\hat{y}^1, \hat{t}^1), \dots, (\hat{y}^k, \hat{t}^k)\}$ is not interim efficient in the k-type model relative to beliefs

$$\left\{ \frac{\Pi^1}{1 - \Pi^{k+1}}, \dots, \frac{\Pi^k}{1 - \Pi^{k+1}} \right\},$$

then there exists an incentive compatible allocation $\{(\bar{y}^1, \bar{t}^1), \dots, (\bar{y}^k, \bar{t}^k)\}$ that Pareto dominates $\{(\hat{y}^1, \hat{t}^1), \dots, (\hat{y}^k, \hat{t}^k)\}$ and satisfies

$$\sum_{i=1}^{k} \frac{\Pi^i}{1 - \Pi^{k+1}} \left[U^i(\bar{y}^i, \bar{t}^i) - U^i(\hat{y}^i, \hat{t}^i) \right] \geqslant 0.$$

Consider the solution to

$$\max V^{k+1}(y^{k+1}, t^{k+1}) \qquad \text{subject to}$$

$$(\mathrm{IC}^{k+1}) \qquad V^k(\bar{y}^k, \bar{t}^k) \geqslant V^k(y^{k+1}, t^{k+1})$$

and

$$(\mathrm{IR}_0^{k+1}) \qquad U^{k+1}(y^{k+1}, t^{k+1}) \geqslant -\sum_{i=1}^{k} \frac{\Pi^i}{\Pi^{k+1}} \left(U^i(\bar{y}^i, \bar{t}^i) - U_0^i \right) + U_0^{k+1}.$$

Note that relative to Program III^{k+1}, both the (IC^{k+1}) and the (IR_0^{k+1}) constraints are relaxed. Hence if $(\bar{y}^{k+1}, \bar{t}^{k+1})$ is a solution to the above program, $V^{k+1}(\bar{y}^{k+1}, \bar{t}^{k+1}) \geqslant V^{k+1}(\hat{y}^{k+1}, \hat{t}^{k+1})$, and so ${}^{k+1}\bar{\mu}^{\cdot} \equiv \{(y^{-i}, t^{-i})\}_{i=1}^{k+1}$ Pareto dominates ${}^{k+1}\hat{\mu}^{\cdot}$. Because ${}^{k+1}\bar{\mu}$ is incentive compatible (from the proof of Proposition 2) and individually rational relative to $(\Pi^1, \dots, \Pi^{k+1})$, it contradicts the interim efficiency of ${}^{k+1}\hat{\mu}^{\cdot}$. *Q.E.D.*

APPENDIX C: Proof of Proposition 8*

Suppose that $\mu_0^{\cdot}$ is an incentive compatible allocation and let $\Pi^{\cdot}$ be the agent's prior beliefs. Choose $\mathring{\Pi}^{\cdot} \neq \Pi^{\cdot}$ (such that $\mathring{\Pi}^i > 0$ for all i) and a WIE allocation $\mathring{\mu}^{\cdot}$ such that for all i, $V^i(\mu_0^i) > V^i(\mathring{\mu}^i)$. Also choose an outcome μ_* such that for all i, $V^i(\mathring{\mu}^i) > V^i(\mu_*)$, $U^i(\mu_0^i) > U^i(\mu_*)$ and $U^i(\mathring{\mu}^i) > U^i(\mu_*)$. Consider the mechanism m_0 in which the principal and agent simultaneously make announcements. The principal announces a number in the set $\{1, \ldots, n\}$ and a letter in the set $\{a, b\}$. The agent announces a letter in the set $\{a, b\}$. If both letters announced are "a" and the principal announces "i", then the contract calls for outcome μ_0^i. If both letters are "b" and the principal announces "i", then $\mathring{\mu}^i$ is the outcome. If the letters differ, the outcome is μ_*.

Notice that one equilibrium of the mechanism m_0 is for both players to announce "a" and for the principal to announce her true type. Another equilibrium is for both players to announce "b" and for the principal to announce her true type.

We claim that, if m_0 is the initial contract, there exists an equilibrium of the three-stage game in which all types of principal propose m_0, these proposals are accepted, and the resulting allocation is $\mu_0^{\cdot}$. In constructing this equilibrium, choose the "b-announcement" equilibrium as the continuation equilibrium in m_0 when the principal proposes a contract other than m_0 and the agent rejects this proposal.

This continuation equilibrium results in allocation $\mathring{\mu}^{\cdot}$. Using the methods of the proof of Theorem 1, we can show that for any mechanism $\tilde{m}$, there exist beliefs $\tilde{\Pi}^{\cdot}$ and an associated equilibrium of the continuation game beginning in the second stage in which, for all i, the type i principal's payoff is no greater than $V^i(\mathring{\mu}^i)$ and hence less than $V^i(\mu_0^i)$. Choose this continuation equilibrium if $\tilde{m} \neq m_0$. If the principal proposes m_0, then choose the "a-announcement" equilibrium in the third-stage continuation equilibrium, whether or not the agent accepts the proposal. The agent's beliefs after such a proposal are just the prior beliefs $\Pi^{\cdot}$. One can verify that this indeed constitutes the equilibrium we claimed. *Q.E.D.*

APPENDIX D: Bilateral Asymmetric Information and the Quasi-linear Case

We state formal sufficient conditions for Assumption Q to be satisfied, and therefore for the analysis of this paper to carry over to bilateral asymmetric information.

Assumption S: (i) *ϕ_j^i does not depend on j (so we drop the subscript j)*; (ii) *y is one-dimensional and y and t can be any real number*; (iii) *ψ_j^i and $-\phi^i$ are increasing in i, ψ_j^i is increasing in j, there exists $\varepsilon > 0$ such that $d\phi^i/dy > \varepsilon$*; (iv) *for all numbers $\bar{u}$ and $\bar{v}$ there exists a finite solution to*

$$\max_{(y_\cdot, t_\cdot)} \sum_j p_j V^i(y_j, t_j) \qquad \textit{subject to}$$

$$\bar{v} \geqslant \sum_j p_j V^{i-1}(y_j, t_j),$$

$$U_2^i(y_2, t_2) \geqslant U_2^i(y_1, t_1), \quad \textit{and}$$

$$U_1^i(y_1, t_1) \geqslant \bar{u},$$

where the first constraint only applies to $i = 2$; (v) (*sorting*) *$d\psi_j^i/dy$ is increasing in i and j; $d\phi^i/dy$ is decreasing in i*; (vi) *$d^2\phi^i/dy^2$ is increasing in i and $d^2\psi_j^i/dy$ is increasing in j.*

We have a generalization of Proposition 2:

Proposition 2*: *Suppose that (a) $n = m = 2$, (b) preferences are quasi-linear, (c) Assumption S is satisfied, (d) $U_j^i(\mu_{j0}^i)$ is nonincreasing in i for $j = 1, 2$, and (e) $U_2^i(\mu_{20}^i) = U_2^i(\mu_{10}^i)$ for all i. The deterministic RSW* allocation relative to $\mu_{\cdot 0}^{\cdot}$ is the least-cost-separating allocation obtained by solving the following program*:

$$\max_{\{y_\cdot^\cdot, t_\cdot^\cdot\}} \sum w^i\left(\sum\left(t_j^i - \phi^i\left(y_j^i\right)\right)\right) \qquad \textit{subject to}$$

$$(\mathrm{PIC}^1) \qquad \sum_j p_j\left(t_j^1 - \phi^1\left(y_j^1\right)\right) \geqslant \sum_j p_j\left(t_j^2 - \phi^1\left(y_j^2\right)\right),$$

$$(\mathrm{AIC}_2^i) \qquad \psi_2^i\left(y_2^i\right) - t_2^i \geqslant \psi_2^i\left(y_1^i\right) - t_1^i \qquad \textit{for all } i,$$

and

$$(\mathrm{IR}_1^i) \qquad \psi_1^i(y_1^i) - t_1^i \geqslant U_1^i(\mu_{10}^i) \qquad \textit{for all } i.$$

Moreover, for all i, the constraints (AIC_2^i) *and* (IR_1^i) *are binding and the ratio of their (strictly positive) Lagrange multipliers is* $1/p_2$. *Hence Assumption Q is satisfied.*

PROOF: See Maskin-Tirole (1990b).

COROLLARY: Under assumptions (a) through (e) of Proposition 2*, an RSW* allocation is IE* for some strictly positive beliefs, and Theorem 1* applies. *Q.E.D.*

REFERENCES

AGHION, P., AND P. BOLTON (1987): "Contracts as a Barrier to Entry," *American Economic Review*, 77, 388–401.

CHO, I. K., AND D. KREPS (1987): "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102, 179–221.

CHO, I. K., AND J. SOBEL (1987): "Strategic Stability and Uniqueness in Signaling Games," Mimeo, University of Chicago.

DEWATRIPONT, M. (1986): "Renegotiation and Information Revelation Over Time in Optimal Labor Contracts," Chapter 1 of *On the Theory of Commitment with Applications to the Labor Market*, Ph.D. Thesis, Harvard University.

DEWATRIPONT, M., AND E. MASKIN (1989): "Multidimensional Screening, Observability and Contract Renegotiation," Mimeo, Harvard University.

ENGERS, M. (1987): "Signaling with Many Signals," *Econometrica*, 55, 663–674.

FARRELL, J. (1985): "Credible Neologisms in Games of Communication," Mimeo, MIT.

FUDENBERG, D., AND J. TIROLE (1990): "Moral Hazard and Renegotiation in Agency," *Econometrica*, 58, 1279–1320.

GALLINI, N., AND B. WRIGHT (1987): "Technology Licensing under Asymmetric Information," Mimeo, University of Toronto.

GERTNER, R., R. GIBBONS, AND D. SCHARFSTEIN (1988): "Simultaneous Signaling to the Capital and Product Markets," *Rand Journal of Economics*, 19, 173–190.

GROSSMAN, S., AND M. PERRY (1986): "Perfect Sequential Equilibrium," *Journal of Economic Theory*, 39, 97–119.

HART, O., AND J. TIROLE (1988): "Contract Renegotiation and Coasian Dynamics," *Review of Economic Studies*, 105, 509–540.

HELLWIG, M. (1986): "Some Recent Developments in the Theory of Competition with Adverse Selection," Mimeo.

HOLMSTRÖM, B., AND R. MYERSON (1983): "Efficient and Durable Decision Rules with Incomplete Information," *Econometrica*, 51, 1799–1820.

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

LAFFONT, J.-J., AND J. TIROLE (1990): "Adverse Selection and Renegotiation in Procurement," *Review of Economic Studies*, 57, 597–626.

MADRIGAL, V., AND T. TAN (1986): "Signaling and Competition," Mimeo, University of Chicago.

MASKIN, E., AND J. RILEY (1984): "Monopoly with Incomplete Information," *Rand Journal of Economics*, 15, 171–196.

MASKIN, E., AND J. MOORE (1987): "Implementation and Renegotiation," Mimeo, London School of Economics.

MASKIN, E., AND J. TIROLE (1990a): "The Principal-Agent Relationship with an Informed Principal, I: Private Values," *Econometrica*, 58, 379–410.

——— (1990b): "The Principal-Agent Relationship with an Informed Principal, II: Common Values," Mimeo, Harvard University and MIT.

MILGROM, P., AND N. STOKEY (1982): "Information, Trade and Common Knowledge," *Journal of Economic Theory*, 26, 17–27.

MIRRLEES, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38, 175–208.

MYERSON, R. (1983): "Mechanism Design by an Informed Principal," *Econometrica*, 51, 1767–1798.

NOSAL, E. (1988): "Implementing Ex-Ante Contracts," Mimeo, University of Waterloo.

RAMEY, G. (1988): "Intuitive Signaling Equilibria with Multiple Signals and a Continuum of Types," Mimeo, University of California, San Diego.

RILEY, J. (1979): "Informational Equilibrium," *Econometrica*, 47, 331–360.

ROTHSCHILD, M., AND J. STIGLITZ (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics*, 90, 629–650.

RUBINSTEIN, A. (1982): "Perfect Equilibrium in a Bargaining Model," *Econometrica*, 50, 97–110.

SPENCE, A. M. (1974): *Market Signaling*. Cambridge: Harvard University Press.

STIGLITZ, J., AND A. WEISS (1983): "Sorting Out the Differences Between Screening and Signaling Models," Mimeo, Columbia University.

STOUGHTON, N., AND E. TALMOR (1990): "Screening vs. Signalling in Transfer Pricing," Mimeo, UC-Irvine and Tel Aviv University.

WILSON, C. (1977): "A Model of Insurance Markets with Incomplete Information," *Journal of Economic Theory*, 16, 167–207.