



**The Principal-Agent Relationship with an Informed Principal: The Case of Private Values**

Eric Maskin; Jean Tirole

*Econometrica*, Vol. 58, No. 2 (Mar., 1990), 379-409.

Stable URL:

<http://links.jstor.org/sici?sici=0012-9682%28199003%2958%3A2%3C379%3ATPRWAI%3E2.0.CO%3B2-C>

*Econometrica* is currently published by The Econometric Society.

---

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

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/econosoc.html>.

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

---

For more information on JSTOR contact [jstor-info@umich.edu](mailto:jstor-info@umich.edu).

©2003 JSTOR

## THE PRINCIPAL-AGENT RELATIONSHIP WITH AN INFORMED PRINCIPAL: THE CASE OF PRIVATE VALUES<sup>1</sup>

BY ERIC MASKIN AND JEAN TIROLE

We analyze the principal-agent relationship when the principal has private information as a three-stage noncooperative game: contract proposal, acceptance/refusal, and contract execution. We assume that the information does not directly affect the agent's payoff (private values). Equilibrium exists and is generically locally unique. Moreover, it is Pareto optimal for the different types of principal. The principal generically does strictly better than when the agent knows her information. Equilibrium allocations are the Walrasian equilibria of an "economy" where the traders are different types of principal and "exchange" the slack on the agent's individual rationality and incentive compatibility constraints.

**KEYWORDS:** Contract, principal-agent relationship, information revelation, general equilibrium, sequential games of incomplete information.

### 1. INTRODUCTION

THE DEVELOPMENT OF THE THEORY of screening (also called the theory of adverse selection or discrimination) represents a major accomplishment of the economics of information in the last two decades. This theory is often cast in a framework with two parties, a principal and an agent. The principal offers a contract, which the agent decides to accept or reject. The agent has private information about some parameter of his utility function. This parameter determines his "type." The parameter affects the principal's payoff at least indirectly, since the agent's type establishes the class of contracts that he will accept. The literature has developed this model both in the abstract (see Laffont-Maskin (1982) and Guesnerie-Laffont (1984) for unified treatments) and as applied to a variety of interesting economic problems, e.g., labor contracts, optimal taxation, price and quality discrimination, insurance contracts, educational screening, auctions, public goods, and regulation of monopoly.

An important hypothesis of the usual model is that the principal is "uninformed," i.e., does not possess private information when contracting. Thus, the asymmetry of information is one-sided. One can think of many circumstances, however, where such an assumption is too restrictive. For example, in the literature on public good mechanisms (see Green-Laffont (1979) for a comprehensive bibliography) the informational deficiency usually emphasized is the government's (principal's) lack of knowledge of consumers' (agents') preferences. But at the time the government institutes a mechanism for eliciting those preferences, it may well know more than consumers about the cost of supplying

<sup>1</sup> This research was supported by the Sloan Foundation, the U.K. Social Science Research Council, and the National Science Foundation. We are grateful to Drew Fudenberg, David Kreps, Roger Myerson, and a referee for helpful comments. We are much indebted to Andreu Mas-Colell for considerable technical advice, and to a second referee for several important insights. This paper was formerly entitled "Principals with Private Information, I: Independent Values."

the good. Alternatively, when a principal such as the Department of Defense deals with contractors (agents) to develop a missile, it may well have special knowledge of the weapon's strategic value. Similarly, a regulator may have private information about the demand for the regulated good when devising a regulatory scheme for a firm it controls. As for private sector examples, a monopolist might have exclusive information about the quality of the good it sells when offering a warranty/maintenance contract to its customers. Similarly, a manufacturer that proposes a franchising agreement to a new retailer could well have superior data about the state of demand.

In keeping with the bulk of the literature, we restrict attention in this paper and its sequel to the case where only one party has a hand in designing the contract. Following standard terminology, we designate the "principal" as the contract "designer" (proposer) and refer to the party that accepts or refuses the contract as the "agent." We deviate from convention, however, by assuming that the principal, as well as the agent, has private information. This assumption complicates contracting because by her very proposal, the principal may reveal some of what she knows.

The revelation of information by contract proposal was emphasized by Myerson (1983) in his seminal article (see also Crawford (1985)). Myerson and Crawford, however, studied the principal-agent relationship using techniques drawn from cooperative game theory. In particular, Myerson was especially concerned with establishing the nonemptiness of the core.

The purpose of our project, by contrast, is to develop a noncooperative theory of the principal-agent relationship when the principal has private information. Section 2 lays out the "principal-agent" or "contract proposal" game, which comprises three stages. Two parties meet after having learned their private information (type). In the first stage, one party, the principal, proposes a contract. The contract is itself a game form in which each party is given a finite set of messages from which to pick and that specifies an action (e.g., producing some level of output) to be taken by the agent and a transfer from the principal to the agent for each pair of messages chosen by the two parties. The agent accepts or refuses the contract in the second stage. If he refuses, the game is over. If he accepts, players proceed to the third stage, where they carry out the contract, i.e., choose their messages and implement the corresponding action and transfer. We assume that actions and transfers are observable (and verifiable), thereby ruling out any moral hazard. Notice that our framework is the same as the classic screening model,<sup>2</sup> except that the principal has private information

<sup>2</sup> By the "classic screening" model, we mean a model of asymmetric information where the "contract" is either chosen by a player without private information (as, for example, in optimal income taxation (c.f., Mirrlees (1971)), monopolistic nonlinear pricing (c.f., Mussa-Rosen (1978), and Maskin-Riley (1984)), and optimal regulation (c.f., Baron-Myerson (1982) and Laffont-Tirole (1986)), or else is negotiated before the asymmetries arise (c.f., Grossman-Hart (1981)). The standard "signaling" model, by contrast, entails an informed party proposing the contract and an uninformed party accepting or rejecting it (c.f., Spence (1974)).

when contracting.<sup>3</sup> This information, however, is important since it will typically affect the ultimate outcome.

In this project we distinguish between the cases where the agent cares and those where he does not care “directly” about the principal’s type or information. The former case is that of *common values*. In the latter case, where we say that values are *private*, the agent’s expected payoff is a function only of the principal’s behavior, not of her information. Formally speaking, privateness means that, holding the principal’s behavior fixed, her information parameter is an argument neither of the agent’s von Neumann-Morgenstern utility function nor of the probabilities he assigns to the variables entering his utility function.<sup>4</sup>

Of course, even in this case, the agent typically cares *indirectly* about the principal’s information because the outcome of the third stage (i.e., the determination of an action and transfer) may depend on the principal’s message, which in turn is influenced by her information.

In this paper we shall deal exclusively with private values. This hypothesis seems a good approximation for the public good, procurement, and regulation examples mentioned above. For instance, the weapons contractor cares only about its profit and not *per se* about the defense value of the missiles it creates. The reader may wish to keep these three examples in mind as paradigms of the sort of situation we are trying to model. By contrast, the common-values case, where the agent’s payoff depends directly on the principal’s type, is illustrated by our monopoly and franchising examples. Specifically, the consumer of a particular good is ordinarily concerned directly about the quality of that good. We take up this case in Maskin-Tirole (1988).

Section 3 demonstrates that the equilibrium contract in our model generally differs considerably from that of the standard principal-agent framework (where only the agent has private information). Indeed, the principal profits from the agent’s incomplete information about her type. To see why this is so, note that when the principal proposes a contract, she does so subject to two kinds of constraints. There is the requirement that the contract should not leave the agent worse off than with no contract, i.e., the individual rationality (IR) constraint. There are also constraints ensuring that, when the contract is carried out, the agent behaves in the appropriate way given his private information. These are the incentive compatibility (IC) constraints. Now, when the agent knows the value of

<sup>3</sup> In this respect, our framework is a synthesis of the screening and signaling models. In this paper, however, we concentrate on the case of “private values,” whereas the signaling literature to date has concentrated primarily on common values (see the next several paragraphs for the distinction).

<sup>4</sup> To see that, in general, we have to rule out the parameter’s affecting the agent’s probabilities (as well as entering his utility function), think of a model of a moral hazard in which output depends stochastically on the agent’s (unobservable) effort and where the principal has private information about this stochastic relationship. Conventionally, the agent’s utility function does not depend on the principal’s information, but rather on his effort and monetary reward. Nonetheless, the principal’s information *does* directly affect the agent’s probabilistic beliefs about output and, hence, his monetary transfer. Thus the agent’s expected payoff is, after all, a function of the principal’s type. We conclude that such a model is an instance of common values.

the principal's information parameter (the case of "full information"), the IR and IC constraints must hold individually for each type of principal. With incomplete information, however, they need hold only in expectation over the principal's types. Thus, a given type of principal can raise her utility above the full-information level (where all the constraints must be satisfied) by violating some constraints, as long as these violations are offset by the other types. In fact, we can think of different types of principal as trading "slack" with one another: one type, say, accepts some slack on the IR constraint in exchange for being allowed to violate an IC constraint, whereas another type does just the opposite. As we show in Section 3A, generically (in the space of utility functions) there exists a contract in which all types of principal do strictly better than in the case of full information (Proposition 1).

This result depends crucially on the private-values assumption. Consider, by contrast, a Spencian labor market (c.f., Spence (1974)) in which the "principal" is an employee of either high or low productivity. In this case the agent's (employer's) payoff certainly depends directly on the principal's type. It is clear, moreover, that the high productivity employee is likely to be *hurt* by the employer's incomplete information: either she will find herself "pooled" with her low productivity counterpart (in which case her wage will fall short of her marginal product) or else she will have to undertake costly signaling activity (e.g., education) to distinguish herself.<sup>5</sup> Thus in a common-values model, unlike one with private values, there is a conflict among the different types of principal.

We can say much more about the equilibrium of our three-stage game than merely that the different types of principal do better than under full information. Indeed, to continue the trading analogy introduced above, consider the fictitious pure-exchange economy in which the traders are the different types of principal and the goods exchanged are the slack variables. A trader's initial endowment consists of the values of these slack variables under full information (i.e., zero). For reasons exactly paralleling the usual competitive analysis, a Walrasian equilibrium always exists for this economy (Proposition 2). Strikingly, moreover, the Walrasian allocations are *precisely* the perfect Bayesian equilibrium outcomes of our three-stage game (Propositions 6 and 9).

From this Walrasian characterization, we can readily establish the generic local uniqueness (Proposition 10) and Pareto optimality (Pareto optimality is, of course, constrained by the fact that the agent also has private information) of equilibrium. Indeed, a strong concept of Pareto optimality offers an alternative characterization of equilibrium. For a contract to be feasible it must satisfy the IR and IC constraints in expectation. The "expectation," of course, depends on the agent's beliefs about the principal. A feasible contract is *strongly* Pareto optimal (from the point of view of the different types of principal) for given

<sup>5</sup> Such an outcome is impossible with private values. A given type of principal can always simply propose the contract that would obtain if the agent knew her type (the "full information" contract), and the agent will accept regardless of his beliefs. In the labor-market example, by contrast, the employer would reject the full-information contract proposal of the high-productivity employee if he thought there was a chance the employee had low productivity.

beliefs if there exists no other feasible contract, *even for different beliefs*, that Pareto dominates it. A Walrasian allocation is strongly Pareto optimal (Proposition 3, our analog of the First Fundamental Theorem of Welfare Economics). Moreover, every strong Pareto optimum is Walrasian (Proposition 4, the Second Welfare Theorem). Hence, given the above-mentioned equivalence between Walrasian equilibria and the PBE's of our game, the same equivalence holds between the strong Pareto optima and the PBE's (Proposition 7).<sup>6</sup>

To reap the gain from the agent's incomplete information, when values are private, the principal must refrain from revealing her type at the contract proposal stage (otherwise, the IR and IC constraints must hold for that type, rather than just in expectation). To accomplish this concealment, the various types of principal have to "pool," i.e., propose the *same* contract in equilibrium.<sup>7</sup>

After our main analysis, we consider in Section 4 the special, but often-studied case where the principal and agent have quasi-linear objective functions (utilities that are additively separable and linear in transfers). In this nongeneric case, the Walrasian equilibrium of the fictitious economy involves no trade and the (unique) equilibrium outcome of our contract proposal game coincides with that of the standard principal-agent model. In other words, with quasi-linear preferences, the principal neither gains nor loses if her information is revealed to the agent before contracting. Section 5 concludes.

## 2. THE MODEL

We now describe the model. In the conclusion, we argue that several of our simplifying assumptions can be relaxed without affecting the results.

### A. Objective Functions and Information

There are two parties, a principal and an agent. The principal has a von Neumann-Morgenstern utility function  $V(y, t, \alpha)$ , where  $y$  is an observable (and verifiable<sup>8</sup>) action,  $t$  is a monetary transfer (which can assume negative as well as positive values) from the principal to the agent, and  $\alpha$  is a parameter representing the principal's private information or "type." We shall suppose that  $y$ ,  $t$ , and  $\alpha$

<sup>6</sup> This result together with Proposition 10 suggests how powerful the concept of strong Pareto optimality is: there is a continuum of ordinary Pareto optima but, generically, only finitely many strong Pareto optima corresponding to the prior beliefs.

<sup>7</sup> The mere observation that there exists a pooling equilibrium of the contract proposal game is, by itself, a *triviality* and holds irrespective of whether values are private or common. Indeed, it is just a reflection of the "Inscrutability Principle" of Myerson (1983), which notes that *any* possible equilibrium outcome arises from some pooling equilibrium if the set of available contracts is sufficiently large. The real substance of our pooling result is that such a separating equilibrium is *not* possible when values are private and the principal also has private information. Generically, in this latter case, *all* equilibria entail some pooling. In particular, no subset of types of principal is completely separated. (If some subset *were* completely separated, its members would not trade at all with the complementary subset in Walrasian equilibrium, which is generically impossible.)

<sup>8</sup> By "verifiable" we mean that the action is observable by a third party; thus, it can be specified by an enforceable contract.

are real numbers. In the case where the principal is a buyer of some good and the agent is a seller, one can think of  $y$  as the quantity of good delivered to the principal. Nothing turns, however, on whether the action is, in fact, taken by the agent or the principal, since it is observed by both and can be specified by a contract. The function  $V$  increases with  $y$  and decreases with  $t$ . It is continuously differentiable and concave in the pair  $(y, t)$  and strictly concave in  $y$ .

The agent has a von Neumann-Morgenstern utility function  $U(y, t, \theta)$ , where the information parameter  $\theta$  (a scalar) is the agent's type. That  $U$  does not depend on  $\alpha$  embodies the assumption of private values and is an important assumption. (In contrast, our results would be unaffected if  $V$  depended on  $\theta$ . See the conclusion.)  $U$  decreases with  $y$  and increases with  $t$ ; it is continuously differentiable and concave in  $(y, t)$  and strictly concave in  $y$ . We will also assume that it decreases with  $\theta$ :

$$\text{if } \theta_1 < \theta_2, \text{ then } U(y, t, \theta_1) > U(y, t, \theta_2) \text{ for all } (y, t).$$

We shall suppose that in the absence of a contract with the principal, a "null" contract takes effect in which the agent obtains reservation utility  $\bar{u}$ .<sup>9</sup> Throughout the paper, superscripts (indexed by  $i$ ) and feminine pronouns refer to the principal, whereas subscripts (indexed by  $j$ ) and masculine pronouns apply to the agent.

To guarantee the existence of equilibrium, we assume that the feasible actions and transfers lie in compact and convex sets. Let  $\mu$  denote a probability measure on these sets. If, for example,  $\mu$  is discrete,  $\mu(\{y, t\})$  represents the probability of action  $y$  and transfer  $t$ . We will allow contracts to specify a measure  $\mu$  as an outcome. We thus permit random outcomes.

We assume that the parameters  $\alpha$  and  $\theta$  are drawn from known and statistically independent distributions. Parameter  $\alpha$  is known only to the principal, and  $\theta$  only to the agent. We suppose that each parameter can assume only finitely many values:  $\alpha = \alpha^1, \dots, \alpha^n$  with positive probabilities  $\pi^1, \dots, \pi^n$  such that  $\sum_{i=1}^n \pi^i = 1$ , and  $\theta = \theta_1$  and  $\theta_2$  with positive probabilities  $p_1$  and  $p_2$  ( $p_1 + p_2 = 1$ ).

The restriction of the agent's parameter to two values is not essential. It ensures that only a single incentive compatibility constraint is binding (see Lemma 1), which is notationally and expositionally convenient. As the reader can readily check, however, all our results require only that *at least* two (IR or IC) constraints be binding.

To simplify the notation, we define

$$V^i(\mu) \equiv \int V(y, t, \alpha^i) d\mu(\{y, t\}) \quad (i = 1, \dots, n)$$

and

$$U_j(\mu) \equiv \int U(y, t, \theta_j) d\mu(\{y, t\}) \quad (j = 1, 2).$$

<sup>9</sup> Thus, we can assume that the principal and agent *always* sign a contract, since the absence of a contract is just a special case of having one. Formally, we must assume that the null  $(y, t)$  pair (ordinarily  $y = t = 0$ ) belongs to the feasible set.

Let  $t$  be the smallest transfer from the principal to the agent (it may well be negative). We will assume that given this transfer the agent is necessarily worse off than without a contract, regardless of his type or the action  $y$ :

ASSUMPTION 1:  $U_j(y, t, \theta_j) < \bar{u}$  for all  $y$  and  $j = 1, 2$ .

We will suppose also that, regardless of the values of  $\theta$  and  $\alpha$ , there exists a feasible action and transfer that both parties prefer to the null contract, i.e., to the absence of a contract.

### 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” or “mechanism” interchangeably). A mechanism  $m$  in  $M$  specifies (i) a set of possible messages that each party can choose and (ii) for each pair of messages  $s^p$  and  $s_a$  chosen simultaneously by the principal and agent, respectively, a corresponding measure  $\mu$  on the set of deterministic allocations  $(y, t)$ .<sup>10</sup> Thus, a mechanism is a game form that selects a (random) outcome conditional on a pair of (payoff-irrelevant) messages.<sup>11</sup> Observe that, because the principal, as well as the agent, can make announcements, she may be able to reveal information at the third stage (see below) as well as at the contract proposal stage.

For the moment we let  $M$  denote the set of *finite* mechanisms (mechanisms where the number of available messages for each party is finite) for simplicity. For technical reasons, we will slightly expand the set of allowable mechanisms in Section 3D.<sup>12</sup>

Notice that the set  $M$  includes the set of *direct revelation* mechanisms, in which both parties simultaneously announce their types (not necessarily truthfully). Hence, in a direct revelation mechanism (DRM),  $(s^p, s_a) = (\hat{\alpha}, \hat{\theta})$ , where a hat denotes an announced value. We will make considerable use of these DRM's by repeatedly invoking the revelation principle for Bayesian games (see Dasgupta-Hammond-Maskin (1979) or Myerson (1979)). 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 an equilibrium of some DRM in which announcements are truthful.

Observe that if  $\alpha$  assumed only one value (i.e., there were no uncertainty about the principal), the revelation principle would imply that the principal could

<sup>10</sup> Because the outcome contingent on  $s^p$  and  $s_a$  can be random, the mechanism can incorporate a correlating device à la Aumann (1974).

<sup>11</sup> For simplicity, we are restricting attention to mechanisms where there is a single round of messages chosen simultaneously. With no change in the arguments, however, we could extend our results, using the revelation principle, to mechanisms with more than one round.

<sup>12</sup> Specifically, we allow a third party, as well as the principal and agent, to send messages in these mechanisms.



restrict attention to mechanisms where she sends no message in the third stage. Because the conventional assumption in the literature is that the principal has no private information, we shall call a mechanism *standard* if it consists simply of the agent's announcing his type (so that principal announces nothing).

In the second stage, the agent accepts or refuses the contract offered by the principal. He obtains his reservation utility if he refuses.<sup>13</sup> If he accepts, the two parties play the proposed mechanism in the third stage (for instance, they announce their types if the mechanism is direct), and the allocation corresponding to their third period moves is implemented.

The principal's strategy in the three-stage game consists of a choice of mechanism and a choice of message ( $s^P$ ) for each mechanism in  $M$ . The agent's strategy consists of (i) the decision to accept or reject the mechanism and (ii) a choice of announcement ( $s_a$ ) in the mechanism. Both the agent's decisions are contingent on the mechanism proposed.

We are interested in the *perfect Bayesian equilibria* of the overall game.<sup>14</sup> In our framework, such an equilibrium is a vector of strategies—one for each type of player (in our model, there are  $n + 2$  types)—and a vector of beliefs about the other player's type<sup>15</sup> at each information set in the game tree such that (i) the strategies are optimal (i.e., at all points in the tree each type is maximizing expected utility given beliefs and the other types' strategies); (ii) beliefs are derived from Bayes' rule given observed behavior and the equilibrium strategies; and (iii) the principal's beliefs about the agent at the end of the first stage remain the prior beliefs (regardless of her proposal) and the agent's beliefs about the principal are the same at the end of the second stage as at the end of the first.<sup>16</sup> Thus, in particular, we assume that the agent updates his beliefs about the principal's type using Bayes' rule, after observing the contract she proposes. 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

<sup>13</sup> In this respect, our model differs from that of the multiperiod bargaining literature (e.g., Admati–Perry (1986), Fudenberg–Tirole (1983), and Sobel–Takahashi (1983), wherein it is typically assumed that the seller (principal) cannot prevent herself from making another offer if the buyer turns her down initially. Note also that there is no conflict between our “pooling” result and the separating equilibria of the noncooperative bargaining literature. A pooling equilibrium does *not* imply that the different types of principal end up with the same allocation; we can think of a contract as a schedule of allocations—one for each type. The different types self-select in the third stage.

<sup>14</sup> Note that because the set of finite mechanisms is itself infinite, so is the strategy space for the principal. Thus standard equilibrium existence lemmas do not apply. Because the second stage continuation equilibria are sequential (see footnote 16), however, standard results ensure that *they* exist and that their corresponding payoffs are upper hemicontinuous with respect to beliefs.

<sup>15</sup> The principal's beliefs are the probabilities that she assigns to  $\theta$ ; the agent's beliefs are the probabilities that he assigns to  $\alpha$ .

<sup>16</sup> Actually, with requirement (iii) (which, in effect, requires that a player's beliefs about the other's type are not affected by his own actions), our definition of perfect Bayesian equilibrium is somewhat stronger than the usual definition. Indeed, as Fudenberg–Tirole (1989) show, conditions (i)–(iii) would imply that the equilibrium is *sequential* if the principal's strategy space were finite.

players can coordinate over these equilibria by means of some public randomizing device<sup>17</sup> such as a coin 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.

We denote by  $\pi = \{\pi^i\}_{i=1}^n$  the agent's prior beliefs about the principal's type and, by  $\hat{\pi} = \{\hat{\pi}^i\}_{i=1}^n$ , his beliefs after the principal has proposed a contract. We shall call these latter probabilities his posterior beliefs. To study the set of equilibrium outcomes, we will use the revelation principle. For given posterior beliefs  $\hat{\pi} = \{\hat{\pi}^i\}_{i=1}^n$  (such that  $\sum_i \hat{\pi}^i = 1$ ), any outcome of the continuation game between the principal and the agent is also the outcome of a direct revelation game in which, in equilibrium, both parties announce their information parameters truthfully and simultaneously. In Sections 3 and 4, we will first construct the strategies along the equilibrium path, and then consider off-the-equilibrium-path proposals  $m$  for given posterior beliefs  $\hat{\pi}$ . Rather than study the equilibrium of the game described by  $m$ , we will instead work with the equivalent direct mechanism  $\mu^*$ , where  $\mu_j^i$  is the (random) outcome implemented if the principal announces type  $i$  and the agent announces type  $j$ .

### C. The Case of Full Information

For reference, we first examine equilibrium when the principal's information is common knowledge, i.e., the agent knows the value of  $\alpha$  before contracting. We call this the full information case (the principal, of course, does not know the agent's type, but, since this feature is maintained throughout the paper, the terminology should not create confusion). This is merely the standard screening set-up (see footnote 2).

Let us assume that  $\alpha = \alpha^i$ . From the revelation principle, we know that the equilibrium allocation can be attained by a standard DRM  $\mu^i$  where, in equilibrium, the agent reveals his type truthfully. The outcome  $\mu_j^i$  ( $j = 1, 2$ ) that the contract specifies when the agent announces type  $j$  must satisfy two types of constraints. First, the agent must be willing to accept the contract. That is, it must satisfy the individual rationality (IR) constraints: for  $j = 1, 2$ ,  $U_j(\mu_j^i) \geq \bar{u}$ . Second, the type  $j$  agent must tell the truth. This gives rise to the incentive compatibility (IC) constraints: for all  $j, k$ :  $U_j(\mu_j^i) \geq U_j(\mu_k^i)$ .

Actually, in an optimal contract (one maximizing the principal's utility), only two of these four constraints are binding. Because the agent's utility decreases with his type, only the IR constraint for an agent of type 2 is required; the other holds automatically. Moreover, this monotonicity of utility implies (see Lemma 1 below) that only the type 1 IC constraint ( $U_1(\mu_1^1) \geq U_1(\mu_2^1)$ ) can be binding. Thus, when her information is common knowledge, a principal of type  $i$  proposes a

<sup>17</sup> The technical reason for allowing public randomization is to ensure that the equilibrium payoff set of the continuation game is convex. Note that this randomization is in addition to that already built into the mechanism (see footnote 10).

contract  $\{\mu_1^i, \mu_2^i\}$  that solves the full information program:

$$(F^i) \quad \text{Max}_{\{\mu_j^i\}} \sum_{j=1}^2 p_j V^i(\mu_j^i) \quad \text{such that}$$

$$\text{IR}^i: U_2(\mu_2^i) \geq \bar{u} \quad (\rho^i),$$

$$\text{IC}^i: U_1(\mu_1^i) \geq U_1(\mu_2^i) \quad (\gamma^i),$$

where  $\rho^i$  and  $\gamma^i$  denote the Lagrange multipliers for the IR and IC constraints, respectively. Because the sets of actions and transfers are compact and the payoff functions are continuous, a solution to program  $(F^i)$  exists. We will denote it by  $\{\bar{\mu}^i, \bar{\rho}^i, \bar{\gamma}^i\}$ . From Assumption 1, it is clear that both constraints are strongly binding (i.e., both  $\bar{\rho}^i$  and  $\bar{\gamma}^i$  are positive). Let  $\bar{v}^i \equiv \sum_j p_j V^i(\bar{\mu}_j^i)$ . We shall refer to  $\bar{v}^i$  as the type  $i$  principal's full-information payoff, as to  $\bar{\mu}^i = (\bar{\mu}_1^i, \dots, \bar{\mu}_n^i)$  as  $\bar{v} = (\bar{v}^1, \dots, \bar{v}^n)$  as the full-information allocation and payoffs, respectively.

It is clear that, regardless of the agent's information about the principal,  $\bar{v}^i$  provides a lower bound for the type  $i$  principal's payoff in our three-stage game. To see this, suppose that she proposed the mechanism  $\bar{\mu}^i$ . Then irrespective of his beliefs, the agent would accept the proposal because, by definition of the mechanism, he could guarantee himself a payoff of at least  $\bar{u}$  by so doing. Moreover, again by definition, he will announce the truth. Thus, by proposing the mechanism  $\bar{\mu}^i$ , the principal ensures herself the payoff  $\bar{v}^i$ .

### 3. PERFECT BAYESIAN AND WALRASIAN EQUILIBRIA

Our goal is a complete characterization of the equilibria of the principal-agent game, but we begin with a simpler problem: studying the contract that would be proposed by a third party who maximized an arbitrary weighted sum of the payoffs of the different types of principal (Section 3A). We first show that, generically, this third party could implement a contract that Pareto dominates the full-information allocation. It accomplishes this by "pooling" the agent's IR and IC constraints over the different types of principal, i.e., by having the constraints hold only in expectation rather than for each single type. This examination leads naturally to a study of the Walrasian equilibria and Pareto optima of the fictitious pure-exchange economy where the traders are the different types of principal and "exchange" the slack variables corresponding to the agent's IR and IC constraints (Section 3B). The relevance of this competitive analysis is demonstrated when we establish that equilibria in the principal-agent game exist (Section 3C) and correspond exactly to the Walrasian allocations of the fictitious economy. Equilibria therefore inherit the Pareto optimality and local uniqueness properties of Walrasian allocations (Section 3D).

#### A. Unconstrained Pareto Optima

As the starting point of the analysis, let us consider the following thought experiment. Suppose that, rather than the principal, a third party proposes the

contract between the principal and the agent. Assume that it acts to maximize a weighted average of the payoffs of the different types of principal, where the weights are nonnegative but arbitrary and fixed beforehand. Suppose, furthermore, that after the contract is proposed and accepted, the principal's type is made publicly known. The device of a third party is meant to rationalize the objective function; if the principal were proposing the contract, she would certainly not do so to maximize an arbitrary weighted average. The fact that the weights are fixed beforehand and do not depend on knowledge the third party has avoids the complicating possibility that the proposal itself may reveal information. Note also that we do not take account of any IR constraints on the part of the principal.<sup>18</sup> Finally, the assumption that the principal's type becomes public knowledge *ex post* eliminates the issue of incentive compatibility for the principal at the third stage. If  $\alpha$  were not public knowledge, then any announcement the principal made at the third stage would reflect her type; her announcement would have to satisfy the IC constraints. Under our hypothesis, by contrast, the contract can make the outcome directly contingent on the principal's type, without having the principal make announcements.

Given these assumptions, the third party solves the program:

$$(UPO) \quad \begin{cases} \text{Max}_{\{\mu_j^i\}} \sum_i w^i \left( \sum_j p_j v^i(\mu_j^i) \right) & \text{such that} \\ \text{IR: } \sum_i \hat{\pi}^i U_2(\mu_2^i) \geq \bar{u} & (\rho), \\ \text{IC: } \sum_i \hat{\pi}^i U_1(\mu_1^i) \geq \sum_i \hat{\pi}^i U_1(\mu_2^i) & (\gamma). \end{cases}$$

We call the solution to this program an “unconstrained Pareto optimum” (the term “unconstrained” refers to the fact that there are no incentive constraints for the principal). Notice the agent's beliefs,  $\hat{\pi}$ , in program (UPO) may differ from his prior  $\pi$ . The number  $w^i$  ( $\sum w^i = 1$ ) is a nonnegative weight for the type  $i$  principal's payoff. The constraints are individual rationality and incentive compatibility requirements, given the agent's beliefs  $\hat{\pi}$ . By omitting the IC constraint for the type 2 agent, we are invoking the familiar result, proved in the Appendix, that this constraint holds automatically at an optimum of (UPO):

LEMMA 1: *The IC constraint for the type 2 agent is not binding at an optimum of (UPO).*

Lemma 1 allows us to simplify the notation by reducing the number of constraints to two: one IR and one IC. One allocation that satisfies these constraints automatically is the full information allocation  $\bar{\mu}$ . (since this allocation

<sup>18</sup> Recall that the type  $i$  principal can guarantee herself a payoff of  $\bar{v}^i$  by proposing  $\bar{\mu}^i$ . Thus if type  $i$  were the proposer (or the third party were acting exclusively on her behalf), an equilibrium would necessarily have to satisfy the IR constraint of ensuring a minimum payoff of  $\bar{v}^i$ . In the framework of our thought experiment, however, no such constraint applies.

tion satisfies  $IR^i$  and  $IC^i$  for each  $i$ ). We next show, however, that generically, the full-information allocation is *not* unconstrained Pareto optimal. This proposition, proved in the Appendix, embodies much of the economic intuition for what follows.

**PROPOSITION 1:** *For a generic choice of utility functions (i.e., for an open and dense set, relative to the  $C^1$  topology, of utility functions satisfying the conditions of Section 2), there is, for any strictly positive posterior beliefs  $\hat{\pi}$  (i.e., beliefs such that  $\hat{\pi}^i > 0$  for all  $i$ ), an allocation that satisfies the IR and IC constraints for the agent with beliefs  $\hat{\pi}$  and that Pareto-dominates the full-information allocation  $\bar{\mu}$ . (from the perspective of the different types of principal).*

The idea behind the proof of Proposition 1 is readily summarized. The full information contract must satisfy the agent's IR and IC constraints for each type  $i$  of principal. If we introduce a small amount of slack  $-r^i$  and  $-c^i$  on these constraints (where  $r^i$  and  $c^i$  can be negative or positive), the type  $i$  principal can attain the payoff

$$v^i = \bar{v}^i + \bar{\rho}^i r^i + \bar{\gamma}^i c^i,$$

where  $\bar{\rho}^i$  and  $\bar{\gamma}^i$  are the shadow prices for the type  $i$  IR and IC constraints, respectively. Now, as long as

$$(1) \quad \sum_i \hat{\pi}^i r^i \leq 0 \quad \text{and} \quad \sum_i \hat{\pi}^i c^i \leq 0,$$

the agent's constraints hold in expectation. Generically, we can choose  $\{r^i, c^i\}_{i=1}^n$  satisfying (1) such that  $v^i - \bar{v}^i$  is nonnegative for all  $i$  and strictly positive for some  $i$ . The allocation,  $\mu_{\bullet}$  corresponding to this choice (i.e., that for all  $i$  solves program  $F^i$  when the constraints are replaced by  $U_2(\mu_2^i) \leq \bar{u} - r^i$  and  $U_1(\mu_1^i) \geq U_1(\mu_2^i) - c^i$ ) Pareto-dominates  $\bar{\mu}$ . Indeed, we can think of  $\mu_{\bullet}$  as being generated by the different types of principal "trading" slack variables. Under this interpretation, the full-information allocation corresponds to autarky.

**REMARK 1:** For there to exist gains from trade, it is clear that at least two constraints for the agent must be binding. Hence, there must be at least two types of agent.

**REMARK 2:** Proposition 1 accords with the observation in Hirshleifer (1971) that the premature disclosure of information may destroy advantageous trading opportunities. It goes well beyond Hirshleifer's analysis, however, by the demonstration that forestalling disclosure makes possible improvements that are *Paretian* for the different realizations of the principal's type (Hirshleifer considers only *ex ante* improvements, i.e., moves that improve welfare in expectation over the various types).

As we will see, equilibria of the three-stage game turn out to be efficient in a sense much stronger than unconstrained Pareto optimality. A UPO allocation is

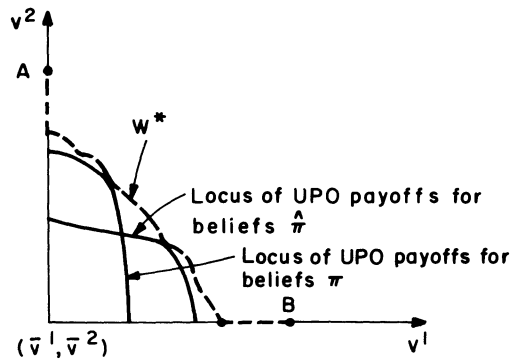
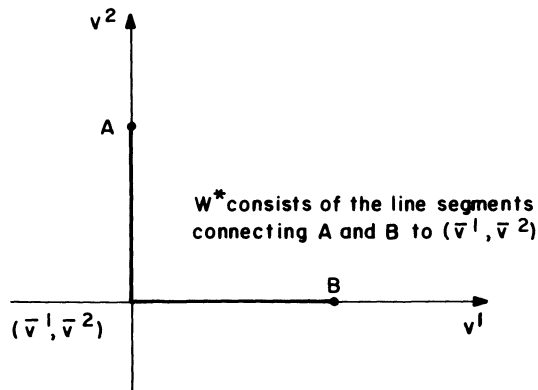


FIGURE 1(a).—The generic case.


 FIGURE 1(b).— $W^*$  in the quasi-linear case.

defined for fixed beliefs, whereas a *strong unconstrained Pareto optimum* (SUPO) allows beliefs themselves to be control variables. (See the Remark following Proposition 5 for an indication of how much stronger SUPO is than UPO.) An allocation  $\mu^*$  is SUPO for beliefs  $\hat{\pi}$  if (i) it is UPO for those beliefs, and (ii) no other UPO allocation  $\hat{\mu}^*$  for any beliefs  $\hat{\pi}$  Pareto-dominates  $\mu^*$ , where the Pareto-domination must be strict (i.e.,  $\sum_j p_j V^i(\hat{\mu}_j^*) > \sum_j p_j V^i(\mu_j^*)$  for all  $i$ ) if  $\hat{\pi}$  is not strictly positive. The set of SUPO payoff vectors is thus given by

$$W^* = \{(v^1, \dots, v^n) \mid \text{there exists a UPO allocation } \mu^* \text{ with beliefs } \hat{\pi} \text{ such that } v^i = \sum_j p_j V^i(\mu_j^*) \text{ for all } i; \text{ moreover, there exist no other UPO allocation } \hat{\mu}^*, \text{ for beliefs } \hat{\pi}, \text{ such that } \sum_j p_j V^i(\hat{\mu}_j^*) \geq \sum_j p_j V^i(\mu_j^*) \text{ for all } i, \text{ with strict inequality for some } i \text{ and where all inequalities are strict if } \hat{\pi} \text{ is not strictly positive}\}.$$

The set  $W^*$  is therefore the “outer envelope” of the UPO payoff loci as beliefs vary. (See Figure 1a. Figure 1b depicts the nongeneric quasi-linear case studied in Section 4.) Notice that the concept of SUPO requires an allocation to be *strictly*

Pareto-dominated (i.e., each type of principal does strictly better in some alternative allocation) to fail as a candidate for an optimum when the beliefs for the alternative allocation are not strictly positive. Thus, it is possible for one SUPO allocation to weakly Pareto-dominate another (see Figure 1a, in which the portions of  $W^*$  that lie along the axes exhibit Pareto dominance).

### B. The Fictitious Competitive Economy

Let  $V_I^i(r^i, c^i)$  denote the type  $i$  principal's indirect utility when there is slack  $-r^i$  and  $-c^i$  in the  $IR^i$  and  $IC^i$  constraints, respectively. Thus  $V_I^i(r^i, c^i)$  is the maximized value of  $(F_\star^i)$ :

$$(F_\star^i) \quad \begin{cases} \text{Max } \sum_j p_j V^i(\mu_j^i) & \text{such that} \\ U_2(\mu_2^i) \geq \bar{u} - r^i, \\ U_1(\mu_1^i) \geq U_1(\mu_2^i) - c^i. \end{cases}$$

Suppose that the type  $i$  principal is allowed to “buy” negative slack (i.e., to sell slack) in the  $IR^i$  and  $IC^i$  constraints at prices  $\rho$  and  $\gamma$ , respectively, subject to the “budget” constraint that the value of the negative slack purchased be nonpositive. She then solves:

$$(D^i) \quad \begin{cases} \text{Max } V_I^i(r^i, c^i) & \text{subject to} \\ \{r^i, c^i\} \\ \rho r^i + \gamma c^i \leq 0. \end{cases}^{19}$$

Let  $D^i(\rho, \gamma)$  denote the type  $i$  principal's “demand correspondence,” i.e., the solution to the program  $(D^i)$ . We thus envision a competitive, pure exchange economy where the traders are the different types of principal and buy and sell slack. Although we have not (yet) restricted the  $(r^i, c^i)$  pairs to a compact set, it is clear that a solution to  $(D^i)$  exists, since the set of feasible pairs  $(y, t)$  is compact. In fact, a solution to  $(D^i)$  must satisfy the budget constraint with equality:

LEMMA 2: If  $(\hat{r}^i, \hat{c}^i) \in D^i(\rho, \gamma)$ , then  $\rho \hat{r}^i + \gamma \hat{c}^i = 0$ .

The proof of Lemma 2 is standard (see Maskin-Tirole (1986)).

A Walrasian equilibrium of this fictitious economy is a pair of positive prices  $(\rho, \gamma)$ , and a choice of negative slack variables<sup>20</sup>  $(r^i, c^i)$  for each type  $i$  such that:

$$(2) \quad \sum_i \pi^i r^i = 0,$$

$$(3) \quad \sum_i \pi^i c^i = 0,$$

<sup>19</sup>Of course, the indirect utility function,  $V_I^i(r^i, c^i)$ , in program  $(D^i)$  already incorporates a maximization over allocations  $\mu$ .

<sup>20</sup>Note that, by referring to  $r^i$  and  $c^i$  as “negative slack” variables, we do not mean that their values are negative. Rather, we are saying only that  $-r^i$  and  $-c^i$  are slack variables.

and

$$(4) \quad (r^i, c^i) \in D^i(\rho, \gamma).$$

Conditions (2) and (3) are “market clearing” requirements, which ensure that the “average” amount of negative slack demanded for each constraint equal the average supply, zero. Condition (4) simply requires that each trader’s choice of slack maximize her (indirect) utility given her budget constraint. We next observe that a Walrasian equilibrium exists in our model for reasons analogous to those in the classical competitive model.

**PROPOSITION 2:** *There exists a Walrasian equilibrium of the fictitious economy relative to any beliefs  $\hat{\pi}$ .*

The proof (which can be found in Maskin-Tirole (1986)) is standard from general equilibrium theory. It suffices to check that the utility functions  $V_j^i(r^i, c^i)$  satisfy the requisite continuity and concavity properties and then to apply the usual Debreu (1959) techniques.

Just as an ordinary Walrasian equilibrium is Pareto efficient, so an equilibrium of our fictitious economy has attractive efficiency properties.

**PROPOSITION 3:** *A Walrasian equilibrium of the fictitious competitive economy is strongly unconstrained Pareto optimal (SUPO).<sup>21</sup>*

Notice that a Walrasian allocation  $\mu^*$  is SUPO even when the corresponding beliefs  $\hat{\pi}$  are degenerate, i.e.,  $\hat{\pi}^i = 1$  for some  $i$ . Now, with such beliefs, the type  $i$  principal’s utility from this allocation is just the full information level,  $\bar{v}^i$ . Therefore, because generically the full information allocation  $\bar{\mu}^*$  is not SUPO, we conclude that (generically) at least one other type of principal does *strictly* better with  $\mu^*$  than with  $\bar{\mu}^*$ .

Propositions 2 and 3 together imply that there is a SUPO allocation for any beliefs  $\hat{\pi}$ . This result relies importantly on the private-values assumption. By contrast, consider the (common-values) labor market example of the introduction. In that model, it is readily checked that there exist “pessimistic” beliefs on the part of the agent (beliefs that assign a comparatively high probability to the “bad” type of principal) relative to which *any* allocation is Pareto dominated by some allocation for more optimistic beliefs. Roughly speaking, this is because, when values are common, the agent suffers from the principal’s type being bad. Thus when the probability of the bad type is high, the agent must be paid correspondingly high compensation (i.e., the principal’s wages are low), implying that the principal’s types have low payoffs. Note that with private values, there is no such thing as pessimistic or optimistic beliefs since the agent does not care about the principal’s type.

<sup>21</sup> Above we described the SUPO locus as the outer envelope of the UPO locus as beliefs vary. Propositions 2 and 3 together imply that for any beliefs there is a corresponding point on that envelope.



Proposition 3, proved in the Appendix, closely resembles the standard line. It is our analog of the First Fundamental Theorem of Welfare Economics. We also can derive a counterpart to the Second Welfare Theorem.

**PROPOSITION 4:** *If  $\hat{\mu}^*$  is a SUPO allocation for strictly positive beliefs  $\hat{\pi}$  (that is, it belongs to the intersection of the SUPO set and the UPO allocations relative to  $\hat{\pi}$ ), then  $\hat{\mu}^*$  is a Walrasian allocation relative to beliefs  $\hat{\pi}$ .*

**PROOF:** Because  $\hat{\mu}^*$  is SUPO, its slack variables and  $\pi = \hat{\pi}$  solve the program:

$$\text{Max}_{\{\pi, r^i, c^i\}} V_I^{i*}(r^{i*}, c^{i*}) \quad \text{such that}$$

$$(5) \quad V_I^i(r^i, c^i) \geq \hat{v}^i \quad \text{for all } i \neq i^*,$$

$$(6) \quad \sum \pi^i r^i \leq 0, \quad \text{and}$$

$$(7) \quad \sum \pi^i c^i \leq 0,$$

where  $i^*$  is a type such that  $w^{i^*} > 0$  in the UPO program that  $\hat{\mu}^*$  solves, and where

$$(8) \quad \hat{v}^i = \sum_j p_j V^i(\hat{\mu}_j^i) \quad \text{for all } i.$$

Let  $\rho$  and  $\gamma$  denote the Lagrange multipliers of (6) and (7). Since  $\pi = \hat{\pi}$  solves the above program and, for all  $i$ ,  $\hat{\pi}^i$  is strictly between 0 and 1, the first-order condition obtained by differentiating the Lagrangian with respect to  $\pi^i$  is:

$$(9) \quad \rho \hat{r}^i + \gamma \hat{c}^i = 0.$$

The first-order conditions with respect to  $r^i$  and  $c^i$  imply

$$(10) \quad \frac{\partial V_I^i}{\partial r^i} \bigg/ \frac{\partial V_I^i}{\partial c^i} = \rho/\gamma \quad \text{for all } i.$$

In view of (9) and (10) and because  $V_I^i(r^i, c^i)$  is concave in  $(r^i, c^i)$ , we infer that  $(\hat{r}^i, \hat{c}^i) \in D^i(\rho, \gamma)$ . We conclude that  $\{(\rho, \gamma), \{(\hat{r}^i, \hat{c}^i)\}_{i=1}^n\}$  is a Walrasian equilibrium relative to beliefs  $\hat{\pi}$ . *Q.E.D.*

**COROLLARY:** *If  $\hat{\mu}^*$  is a SUPO allocation for strictly positive beliefs  $\hat{\pi}$ , it satisfies the principal's individual rationality and incentive constraints:*

$$(PIR) \quad \sum_j p_j V^i(\hat{\mu}_j^i) \geq \bar{v}^i$$

and

$$(PIC) \quad \sum_j p_j V^i(\hat{\mu}_j^i) \geq \sum_j p_j V^i(\hat{\mu}_j^h),$$

for all  $i$  and  $h$ .

**PROOF:** Because, in view of Proposition 4,  $\hat{\mu}^*$  is a Walrasian allocation, it must give the type  $i$  principal at least the utility she obtains from her "initial

endowment,"  $(r^i, c^i) = (0, 0)$ . But  $V_i^i(0, 0) = \bar{v}^i$ . Hence, we obtain (PIR). Since all types of principal have the same endowment, moreover, they can afford to buy each other's equilibrium allocations. Thus (PIC) follows. *Q.E.D.*

**REMARK:** In our definition of  $W^*$ , we did not require that the principal's utilities exceed the full information levels. Nonetheless, the corollary to Proposition 4 demonstrates that this property holds for all points in  $W^*$  corresponding to strictly positive beliefs. The corollary also vindicates our omission of the principal's IC constraints in the definition of  $W^*$ . It can be shown, however, that both Proposition 4 and its corollary are false when beliefs fail to be strictly positive (see Maskin-Tirole (1986)).

Given that a competitive economy is sufficiently smooth, it generically has only finitely many equilibria. For exactly the same reasons, we can draw such a conclusion in our model.

**PROPOSITION 5:** *For an open and dense subset of utility functions (satisfying the conditions of Section 2) there exist only finitely many Walrasian equilibria relative to  $\bar{\pi}$ .*

The proof of Proposition 5 is standard but uses methods of differential topology that are beyond the scope of this paper. We refer the reader to Mas-Colell (1985) for a comprehensive treatment.

**REMARK:** Propositions 4 and 5 together illustrate how much stronger a concept SUPO is than UPO. For fixed beliefs there is a continuum of allocations solving program UPO. However, generically, only finitely many of these are SUPO.

### C. Equilibrium in the Principal-Agent Game

We now use our results for the competitive economy to study perfect Bayesian equilibrium of the principal-agent game. We first demonstrate that one can construct such an equilibrium from a Walrasian allocation.

**PROPOSITION 6:** *There exists a perfect Bayesian equilibrium of the three-stage contract proposal game. More specifically, for any prior beliefs  $\pi$  and any Walrasian allocation for the fictitious economy relative to  $\pi$ , there exists an equilibrium where all types of principal propose the same contract and where the equilibrium outcome is this Walrasian equilibrium allocation.*

**PROOF:** Consider a Walrasian equilibrium  $\{(\hat{p}, \hat{y}), \{\hat{p}^i, \hat{c}^i\}_{i=1}^n\}$  relative to the prior beliefs  $\pi$ . Let  $\hat{\mu}^*$  be the corresponding allocation and let  $\hat{v}$  be the vector of Walrasian payoffs. From Proposition 2, such an equilibrium exists.

We first construct the equilibrium path of our perfect Bayesian equilibrium. Along the path, all types of principal propose the direct revelation mechanism  $\hat{\mu}^*$ . Because the agent can infer nothing from this proposal, he does not modify his

prior beliefs about the principal's information, i.e.,  $\hat{\pi}^i = \pi^i$  for all  $i$ . The agent, irrespective of his type, accepts the contract in the second stage, and so the principal does not revise her prior probabilities  $(p_1, p_2)$ . Finally, both parties announce their types truthfully in the third stage.

To demonstrate that this behavior forms an equilibrium path, we work backwards from the end. We first show that truthful revelation is optimal for both parties in the third stage; next, that it is in the agent's interest to accept the mechanism  $\hat{\mu}$  in the second stage; and, finally, that for any alternative contract proposal in the first stage, there exist posterior beliefs and a corresponding continuation equilibrium in which no type of principal is better off than on the equilibrium path.

Because  $\hat{\mu}$  is a Walrasian allocation relative to beliefs  $\pi$  it satisfies the agent's IC constraints by definition. Hence, if the principal announces the truth in the third stage, the agent will find it worthwhile to do so too if his beliefs are  $\pi$ . From the corollary to Proposition 4,  $\hat{\mu}$  also satisfies the principal's IC constraints when her beliefs about the agent's type are  $(p_1, p_2)$ . Hence, truth-telling forms a Bayesian Nash equilibrium in the third stage, assuming that the parties have maintained their prior beliefs.

Because the agent obtains at least the utility  $\bar{u}$  in the third stage, it is optimal for him, given his prior beliefs, to accept the proposal  $\hat{\mu}$  in the second stage regardless of his type. Hence, the principal will not update her prior beliefs.

It remains to choose off-the-path strategies and beliefs at the first stage that deter the principal from proposing a contract other than  $\mu$ . Suppose that the principal proposes some other, finite mechanism  $m$ . Because this proposal is never made in equilibrium, beliefs  $\hat{\pi}$  are not determined by Bayes' rule. Instead, they can be arbitrary. For each possible vector of beliefs, there is at least one corresponding continuation equilibrium (see Kreps-Wilson (1982)).<sup>22</sup> Let  $V_m$  be the convex hull of the set of continuation equilibrium payoff vectors (for the principal) corresponding to  $m$ . For any posterior beliefs  $\hat{\pi}$ , let  $V_m^e(\hat{\pi})$  be the set of continuation equilibrium payoff vectors for the principal when  $m$  is proposed and beliefs are  $\hat{\pi}$ . If we suppose that, in the case of multiple equilibria, a random public event (e.g., sunspots) makes the selection, then  $V_m^e(\hat{\pi})$  is a convex and compact subset of  $V_m$ .

For payoff vectors  $v \in V_m$  and beliefs  $\hat{\pi}$  define the correspondence

$$(11) \quad (\hat{\pi}, v) \rightarrow \left\{ \tilde{\pi} \mid \tilde{\pi} \in \arg \max_{\pi} \sum_i \sum \pi^i (v^i - \hat{v}^i) \right\} \times V_m^e(\hat{\pi}),$$

where  $\hat{v}^i$  is the type  $i$  principal's Walrasian payoff. Correspondence (11) is closely analogous to the well-known Debreu mapping used to establish existence of competitive equilibrium. The cross product of the belief and payoff sets is compact and convex, and the correspondence is upper hemicontinuous (see

<sup>22</sup> It need not be the case that either or both types of agent accept the proposal  $m$  in a given continuation equilibrium. But, whether or not the proposal is accepted, there is still an equilibrium payoff.

footnotes 15 and 16) and convex-valued (see footnote 17). Hence, it admits a fixed point,  $(\pi_0, v_0)$ .

Assume first that

$$(12) \quad \sum_i \pi_0^i (v_0^i - \hat{v}^i) > 0.$$

Let  $I = \{i | v_0^i > \hat{v}^i\}$  and  $J = \{i | v_0^i \leq \hat{v}^i\}$ .  $J$  is not empty, because  $\hat{v}$  is SUPO. From (12)  $I$  is not empty. Because  $\pi_0$  maximizes  $\sum \pi^i (v_0^i - \hat{v}^i)$ ,  $\pi_0^i = 0$  for  $i \in J$ . Thus, the agent IR and IC constraints can be written as

$$(13) \quad \sum_{i \in I} \pi_0^i r_0^i \leq 0$$

and

$$(14) \quad \sum_{i \in I} \pi_0^i c_0^i \leq 0,$$

where  $r_0^i$  and  $c_0^i$  denote the negative slack variables associated with  $v_0$ .

Now, each type  $i$  in  $I$  prefers  $v_0^i$  to her competitive payoff, which means she cannot afford  $r_0^i$  and  $c_0^i$  at competitive prices  $\hat{p}$  and  $\hat{\gamma}$ :

$$(15) \quad \hat{p}r_0^i + \hat{\gamma}c_0^i > 0, \quad \text{for } i \in I.$$

But (13) through (15) are clearly inconsistent. Hence, (12) is impossible.

We conclude that  $\sum \pi_0^i (v_0^i - \hat{v}^i) \leq 0$ . Thus, for all  $i$ ,  $v_0^i \leq \hat{v}^i$  (otherwise there would exist  $\hat{\pi}$  such that  $\sum \hat{\pi}^i (v_0^i - \hat{v}^i) > 0$ , a contradiction). Because  $v_0 \in V_m^e(\pi_0)$ ,  $\pi_0$  and  $v_0$  constitute beliefs and corresponding continuation equilibrium payoffs that no type of principal prefers to her Walrasian payoff. Hence if we assign  $\pi_0$  and  $v_0$  to  $m$ , no type of principal will deviate from proposing  $\hat{\mu}$ . Q.E.D.

To summarize the construction of the proof, each type of principal proposes the Walrasian allocation  $\hat{\mu}$  as a direct revelation mechanism. In equilibrium, the agent accepts the proposal, and, in the third stage, the two parties announce their types truthfully. Should the principal propose some other mechanism, the agent's beliefs and the continuation equilibrium are chosen so that all types of principal are no better off than with  $\hat{\mu}$ . That this is possible is particularly clear when  $n = 2$ . Suppose that the principal proposes some out-of-equilibrium mechanism  $m$ . If the agent attaches probability 1 to  $\alpha = \alpha^1$ , then the type 1 principal can derive no more utility than  $\bar{v}^1$ , which is clearly less than that which she derives from  $\hat{\mu}^1$ . Similarly, the type 2 principal obtains less utility from  $m$  if the agent believes  $\alpha = \alpha^2$  than from  $\hat{\mu}^2$ . From continuity and because  $\hat{\mu}$  is SUPO, there exist intermediate beliefs where both types are no better off with  $m$  than with  $\hat{\mu}$ . (see Figure 2).

**REMARK:** Proposition 6 is a reflection of the idea that, far from there being a conflict among the different types of principal, they mutually gain from the agent's incomplete information. They take advantage of this incomplete information by revealing no information until their proposal is accepted and then exploiting the fact that the agent's constraints need hold only in expectation.

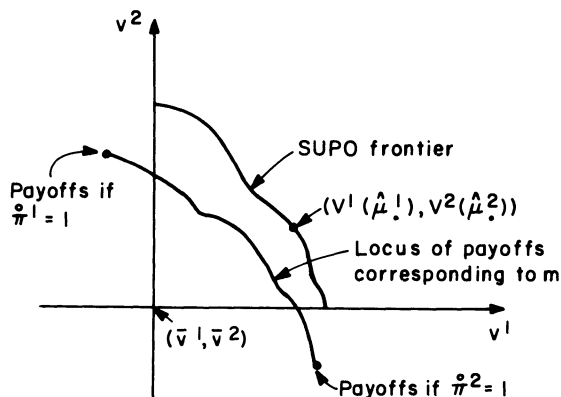


FIGURE 2.—Existence of a PBE corresponding to a Walrasian equilibrium.

#### D. Uniqueness

We proved in subsection 3C that there exists an equilibrium of the three-stage principal-agent game. This equilibrium corresponds to a Walrasian allocation of the fictitious competitive economy. We now investigate uniqueness.

Sequential games of incomplete information are often plagued by a plethora of equilibria. One may wonder whether such is the case here. Can any strong UPO allocation be an equilibrium outcome of the three-stage game? Do there exist any suboptimal equilibria? As we shall see, the answer to both questions is “no.” Indeed, we demonstrate that *only* Walrasian allocations relative to the prior beliefs  $\pi$  can arise as equilibria of the principal-agent game. Since such allocations are, generically, locally unique, the same is, therefore, true of the game’s equilibrium outcomes.

To establish our uniqueness result, we expand the class of permissible mechanisms somewhat. In particular, we now include mechanisms in which a third party, in addition to the principal and agent, chooses from a set of messages. Moreover, rather than just dealing with finite mechanisms, we let the permissible set  $M^*$  include all mechanisms  $m$  such that, if the principal’s beliefs about the agent at the time the mechanism is to be played are given by the prior beliefs  $(p_1, p_2)$ , (a) there exists a perfect Bayesian equilibrium of  $m$  regardless of the agent’s beliefs about  $\alpha$ ; (b) for any SUPO payoff vector,  $(\hat{v}^1, \dots, \hat{v}^n)$ , there exists, for some vector of agent’s beliefs  $\hat{\pi}$ , an equilibrium of  $m$  for which the equilibrium payoffs,  $(v^1, \dots, v^n)$ , satisfy  $v^i \leq \hat{v}^i$  for all  $i$ . Conditions (a) and (b) are admittedly technical but express the natural requirements that (i) the principal should be able to predict the outcome of her proposal (equilibrium should exist) and (ii) equilibrium should be well-behaved as a function of beliefs ((b) is satisfied if the equilibrium payoff correspondence is upper hemicontinuous and convex valued). The conditions, moreover, are automatically satisfied for finite mechanisms. As we noted in the proof of Proposition 6, (a) is guaranteed for finite mechanisms by sequentiality (see Kreps-Wilson (1982)). Condition (b) for finite mechanisms was established in the proof of Proposition 6. Indeed, one can

easily confirm that conditions (a) and (b) were the only properties of mechanisms required for demonstrating the existence of equilibrium in our principal-agent game. Hence, Proposition 6 continues to hold for the larger class  $M^*$ .

Besides enlarging the class of available mechanisms, we also strengthen our assumptions about the agent's utility function. Specifically, we suppose that it satisfies a conventional "sorting" assumption.

ASSUMPTION 2:

$$-\frac{U_y(y, t, \theta_1)}{U_t(y, t, \theta_1)} < -\frac{U_y(y, t, \theta_2)}{U_t(y, t, \theta_2)} \quad \text{for all } (y, t).$$

PROPOSITION 7: *Let  $M^*$  be the class of admissible mechanisms. Then, given Assumption 2 (in addition to the assumptions of Section 2), any perfect Bayesian equilibrium of the principal-agent game is Strong Unconstrained Pareto Optimal (SUPO).*

That equilibrium allocations must be Pareto optimal relies on the ability of the principal to break an inefficient equilibrium by proposing an alternative mechanism that, whatever the agent's beliefs turn out to be, makes (at least) one of her types better off. This "equilibrium breaking" can be accomplished by the following simple mechanism  $m^*$ . First, the principal and agent announce probability vectors  $\pi_p$  and  $\pi_a$  (corresponding to the agent's beliefs about the principal's type when  $m^*$  is proposed). If  $\pi_p \neq \pi_a$ , the null contract is imposed. If  $\pi_p = \pi_a = \pi$ , the principal and agent play the Walrasian direct-revelation game corresponding to  $\pi$ . I.e., they announce their types simultaneously, and the outcome is the Walrasian allocation for the announced types relative to  $\pi$  (the game must be somewhat modified if there are multiple Walrasian equilibria). Notice that it is an equilibrium of this game for the two players to announce the agent's true beliefs and then announce their true types. This equilibrium, therefore, is Walrasian relative to the agent's true beliefs, and so does the trick of equilibrium breaking.

The weakness of  $m^*$  is that, although the above "truthful" equilibrium may be particularly salient, there are other, "perverse" equilibria of  $m^*$  in which the players (a) announce different beliefs, or (b) announce the same but false beliefs, or (c) announce their types falsely. The proof of Proposition 7 (see Appendix) constructs a more elaborate mechanism, based on  $m^*$ , in which these perverse equilibria are eliminated and only the Walrasian outcome remains.

The Pareto optimality of equilibrium depends importantly on private values. As we noted following Proposition 3, SUPO allocations do not even exist relative to all beliefs in common-values models such as the Spencian labor market. Moreover, even for beliefs relative to which a SUPO allocation *does* exist, there can be many inefficient equilibria even if the principal uses the sort of mechanisms invoked in the proof of Proposition 7. This is because to break an inefficient equilibrium, as we have noted, the principal needs to propose a mechanism that, regardless of the agent's beliefs, is better for one of her types.

But with common values and “pessimistic” beliefs by the agent (beliefs that attach high probability to the bad type(s) of principal), all the principal’s types may actually be worse off than in the inefficient equilibrium. Thus the equilibrium cannot be broken.

One can interpret Proposition 7 as an illustration of the idea that if, relative to beliefs, there are gains from trade, the principal ought to be able to exploit them. The common-values model is not a counterexample to this principle because there, if the principal tries to overcome the inefficiency, the agent’s beliefs may change in such a way that there *are* no gains from trade.

**PROPOSITION 8:** *Given the hypotheses of Proposition 7, any perfect Bayesian equilibrium allocation  $\mu^*$  of the three-stage game is a Walrasian allocation relative to prior beliefs  $\pi$ .*

**PROOF:** From Proposition 7,  $\mu^*$  is SUPO. Hence, because it satisfies the IR and IC constraints of program (UPO) for beliefs  $\pi$ , it is SUPO for  $\pi$ . By assumption,  $\pi$  is strictly positive, and so, from Proposition 4,  $\mu^*$  is Walrasian relative to  $\pi$ . *Q.E.D.*

We noted above that any equilibrium allocation can be thought of as arising from a pooling equilibrium, in which all types of principal propose the same mechanism. Proposition 8 demonstrates that, in general, some pooling is *essential* in equilibrium. A Walrasian allocation generically strictly Pareto dominates the full-information payoff vector. Thus the fact that the equilibrium allocation is necessarily Walrasian implies that the principal cannot perfectly reveal her type by her proposal.

We know from Proposition 5 that the Walrasian equilibria of the fictitious economy are generically finite in number. In view of Proposition 8, we can conclude the same for the perfect Bayesian equilibria of our three-stage game.

**PROPOSITION 9:** *For an open and dense set of utility functions (satisfying the hypotheses of Proposition 7), there exist only finitely many perfect Bayesian equilibrium allocations of our principal-agent game.*

Propositions 7 through 9 are obtained by extending the class of mechanisms beyond DRM’s. Another, and quite different route to efficiency and uniqueness is to refine the concept of perfect Bayesian equilibrium. Specifically, even if the principal is constrained to propose only allocations (i.e., DRM’s), the application of the Farrell (1985) Grossman-Perry (1986) (FGP) refinement<sup>23</sup> again rules out

<sup>23</sup> In our context, this refinement requires that there does not exist a subset of types  $S$  and an alternative allocation such that types in  $S$  (weakly) gain (and the other types lose) relative to equilibrium and the allocation satisfies the agent’s IC and IR constraints if the agent updates his prior beliefs so that, ignoring the renormalization needed to have probabilities sum to 1, the posterior probability of a type who strictly gains is the same as the prior probability, that of a type who strictly loses is zero, and that of a type who is indifferent is intermediate.

all but SUPO allocations in equilibrium. To continue our Walrasian metaphor, a rough intuition for this result is that the core coincides with the set of Walrasian allocations (in large economies). Thus, if an allocation is Walrasian, there is no subset of principal's types that can make themselves better off by trading among themselves. Conversely, if an allocation is not Walrasian, there *does* exist such a coalition.

**PROPOSITION 10:** *Walrasian allocations relative to prior beliefs  $\pi$  satisfy the FGP refinement. Conversely, if either  $n = 2$  or there exists a unique Walrasian equilibrium for any  $\pi$ , any perfect Bayesian equilibrium allocation  $\mu^*$  of the three-stage game (where the principal proposes DRM's) that satisfies the FGP refinement is a Walrasian allocation relative to prior beliefs  $\pi$ .*

**PROOF:** See the Appendix.

#### 4. QUASI-LINEAR UTILITIES

Much of the incentives literature concerns the special case of quasi-linear objective functions for the principal and the agent:

$$V^i = \phi^i(y) - t \quad (i = 1, \dots, n),$$

and

$$U_j = t - \psi_j(y) \quad (i = 1, 2).$$

For our purposes, the most important feature of these functions is that the shadow values of the two constraints in the full-information program are *independent* of the principal's type.<sup>24</sup> That is, the marginal rate of substitution between the two slack variables is the same for any type; and so there are no gains to be reaped from trade. The Walrasian equilibrium of the fictitious competitive economy is autarky. Hence, Proposition 1 does not pertain to quasi-linear utilities. Indeed, from previous analysis, we immediately obtain the following proposition.

**PROPOSITION 11:** *In the quasi-linear case, the unique equilibrium payoff vector of the three-stage game is the full-information vector  $\bar{v}$ .*

Proposition 11 asserts that, with quasi-linear utilities, the principal neither gains nor loses if her type is revealed to the agent before the game is played. Of course, this is an outcome of the nongeneric nature of the quasi-linear case.

<sup>24</sup> Because the payoff functions are linear in transfers, we might as well assume that transfers in a solution to  $(F^i)$  are deterministic; we can replace any random transfer by its mean without affecting anything. Now, forming the Lagrangian for  $(F^i)$  and optimizing with respect to the transfers implies that  $\bar{p}^i = 1$  and  $\bar{\gamma}^i = p_1$  regardless of the value of  $i$ .



## 5. SUMMARY

When values are private, the principal strictly gains, in general, by concealing her type until the contract she proposes is carried out. This concealment enables her to be constrained by the agent's individual rationality and incentive compatibility constraints merely in expectation, rather than type by type. One can, in fact, view the different types of principal as competitive traders in the slack variables associated with these constraints; one trader's violation of a constraint is counterbalanced by another trader's accepting some slack. In fact, the equilibria of the three-stage principal-agent game correspond exactly to the Walrasian allocations of this competitive economy (and so, in particular, they are efficient in a strong sense).

The Walrasian interpretation is illuminating in several respects. As we have just indicated, it helps us understand why the principal gains from pooling and how she profits from the agent's ignorance of her type. It also explains why, in equilibrium, the principal's *own* incentive compatibility constraints are not binding. Just as consumers trading from equal endowments do not envy each other's allocations in Walrasian equilibrium, so no type of principal prefers the equilibrium allocation of some other type.

The analogy with Walrasian equilibrium, however, relies on the privateness of values and the absence of moral hazard. We have already noted that, in common value models, inefficient (and hence non-Walrasian) equilibria may exist in large numbers (see Maskin-Tirole (1988) for greater elaboration). In such models, unlike that of this paper, it is no longer true that, without loss, the principal can postpone revealing her type until the third stage. She may wish to disclose information about herself in order to influence the agent's action. Her proposal must therefore strike some balance between total disclosure and complete concealment.

Although our model is already quite general, many of our assumptions can be relaxed further. The two crucial assumptions for our results are that (i) the principal's information parameter does not enter the agent's utility function (thereby avoiding signaling phenomena) and (ii) the full-information program includes at least *two* binding agent constraints (so that the principal's types are able to trade slack variables). Thus our results would not be affected by (a) multidimensional type and action spaces; (b) nonmonotonic utility functions; (c) one-sided common values, in which the principal's utility depends on the agent's information (indeed, in public sector applications, where the principal acts on behalf of society (e.g., the public good or regulation examples), her objective function may take account of the agent's welfare; in that case,  $V$  is a function of  $\theta$  as well as of  $(y, t, \alpha)$ ); (d) statistical dependence between  $\alpha$  and  $\theta$  (in this case, parties' expectations must be made conditional on their own types); (e) arbitrary number of agent's types (focusing on two types allowed us to simplify exposition since only two constraints—one IR and one IC—were binding; what matters is that there be at *least* two binding constraints); (f) reservation utilities that depend on the agent's type (for the same reason).

None of the generalizations requires further argument; they demand only more involved notation.

*Department of Economics, Harvard University, Cambridge, MA 02138, U.S.A.*  
and

*Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.*

*Manuscript received May, 1986; final revision received March, 1989.*

## APPENDIX

**PROOF OF LEMMA 1:** To see that the IC constraint for the type 2 agent holds, consider the program (UPO). We will show that a solution to this program satisfies the deleted constraint. Observe first that, if  $\mu^*$  is a solution to the program (UPO), then

$$(*) \quad \sum_i w^i V^i(\mu_1^*) \geq \sum_i w^i V^i(\mu_2^*).$$

Formula (\*) holds because, were it violated, the (pooling) allocation  $\tilde{\mu}^*$ , defined so that, for all  $i$ ,  $\tilde{\mu}_1^i = \tilde{\mu}_2^i = \mu_2^*$  would satisfy the constraints of the program (UPO) and generate a higher value of the maximand than  $\mu^*$ . Now, if  $\mu^*$  violates the deleted IC constraint (i.e., the type 2 agent strictly prefers  $\mu_1^*$  to  $\mu_2^*$ ), let us define  $\hat{\mu}^*$  so that, for all  $i$ ,  $\hat{\mu}_1^i = \hat{\mu}_2^i = \mu_1^*$ . The allocation  $\hat{\mu}^*$  satisfies the constraints of the program (UPO) and, from (\*), generates at least as high a value of the maximand as  $\mu^*$ . Indeed, because the type 2 agent *strictly* prefers  $\mu_1^*$  to  $\mu_2^*$ , we can slightly decrease the transfer from the principal to the agent in  $\hat{\mu}^*$ . (Assumption 1 guarantees that such an increase is possible) without violating the constraints. But then  $\hat{\mu}^*$  generates a higher value of the maximand than  $\mu^*$ , a violation of  $\mu^*$ 's optimality. *Q.E.D.*

**PROOF OF PROPOSITION 1:** Consider the solution  $(\bar{\mu}', \bar{p}', \bar{\gamma}')$  to  $(F')$ , where  $\bar{p}'$  and  $\bar{\gamma}'$  are the shadow prices of the IR' and IC' constraints. For any two types of principal, say 1 and 2, it can be shown that, for almost all choices of utility functions,  $V^1$  and  $V^2$ , satisfying the Section 2 assumptions, the corresponding shadow prices satisfy  $\bar{p}^1/\bar{\gamma}^1 \neq \bar{p}^2/\bar{\gamma}^2$ .<sup>25</sup>

For an arbitrary (random) allocation  $\mu^i$ , let  $r^i(\mu^i)$  and  $c^i(\mu^i)$  be the negatives of the slack variables associated with the IR' and IC' constraints:

$$r^i(\mu^i) \equiv \bar{u} - U_2(\mu_2^i) \quad \text{and} \quad c^i(\mu^i) \equiv U_1(\mu_2^i) - U_1(\mu_1^i).$$

In particular,  $r^i(\bar{\mu}^i) = 0$  and  $c^i(\bar{\mu}^i) = 0$ . Moreover, for beliefs  $\hat{\pi}$ , the constraints of the (UPO) program can be expressed as

$$IR: \sum_i \hat{\pi}^i r^i(\mu^i) \leq 0 \quad \text{and} \quad IC: \sum_i \hat{\pi}^i c^i(\mu^i) \leq 0.$$

Thus, to satisfy the constraints of (UPO), the negatives of the slack variables need only be nonpositive *on average*, and not for each value of  $i$  individually.

<sup>25</sup> "Almost all" means "for an open and dense subset." Andreu Mas-Colell has provided us with a proof of this assertion.

Consider the following perturbed version of the full-information program ( $F'$ ):

$$(F'_*) \quad \begin{cases} \text{Max } \sum_j p_j V^i(\mu'_j) & \text{such that} \\ U_2(\mu'_2) \geq \bar{u} - r', \\ U_1(\mu'_1) \geq U_1(\mu'_2) - c'. \end{cases}$$

By definition of the shadow prices  $\bar{\rho}'$  and  $\bar{\gamma}'$ , the maximized value,  $v'_*$ , of the maximand approximately equals  $\bar{v}' + \bar{\rho}'r' + \bar{\gamma}'c'$ , for small values of  $r'$  and  $c'$ . Let  $\mu'_{*}$  be the solution to  $F'_*$ .

Choose negative slack variables  $(r^1, c^1)$  for the type 1 principal; define negative slack variables  $\{r^2 = -(\hat{\pi}^1/\hat{\pi}^2)r^1, c^2 = -(\hat{\pi}^1/\hat{\pi}^2)c^1\}$  for the type 2 principal; and take  $\{r' = 0, c' = 0\}$  for types 3 through  $n$ . Via program ( $F'_*$ ), we obtain new maximized utilities for the type 1 and 2 principals (the other types' payoffs are the same as under full information):

$$(a1) \quad v'_* - \bar{v}^1 \approx \bar{\rho}^1 r^1 + \bar{\gamma}^1 c^1,$$

and

$$(a2) \quad v'_* - \bar{v}^2 \approx -(\hat{\pi}^1/\hat{\pi}^2)(\bar{\rho}^2 r^1 - \bar{\gamma}^2 c^1).$$

If  $\bar{\rho}^1/\bar{\gamma}^1 \neq \bar{\rho}^2/\bar{\gamma}^2$  (as is generically the case) one can choose  $(r^1, c^1)$  small enough<sup>26</sup> that

$$(a3) \quad v'_* - \bar{v}^1 > 0 \quad \text{and} \quad v'_* - \bar{v}^2 > 0.$$

For instance, if  $\bar{\rho}^1/\bar{\gamma}^1 > \bar{\rho}^2/\bar{\gamma}^2$ , so that the IR constraint is relatively more costly for the type 1 principal, she can “accept” some slack on the IC constraint in exchange for being permitted “negative slack” on the IR constraint. That is,  $r^1$  is positive and  $c^1$  negative. From our choice of slack variables, the allocation  $\mu'_{*}$  satisfies the constraints of program (UPO). Thus, in view of (a3),  $\bar{\mu}'$  is not unconstrained Pareto optimal. Q.E.D.

**PROOF OF PROPOSITION 3:** Let  $(\rho, \gamma)$  and  $\{(r^i, c^i)\}_{i=1}^n$  be a Walrasian equilibrium relative to beliefs  $\hat{\pi}$  and let  $\mu'_*$  be the corresponding allocation. If this allocation is not SUPO, then there exists a UPO allocation  $\hat{\mu}'_*$  for beliefs  $\hat{\pi}$  such that, for all  $i$ ,

$$(b1) \quad \sum_j p_j V^i(\hat{\mu}'_j) \geq \sum_j p_j V^i(\mu'_j),$$

where inequality (b1) is strict for some  $i$  and, if  $\hat{\pi}$  is not strictly positive, strict for all  $i$ . Let  $\{\hat{r}^i, \hat{c}^i\}_{i=1}^n$  be the negative slack variables associated with the allocation  $\hat{\mu}'_*$ . Then, from the IR and IC constraints of program (UPO),

$$(b2) \quad \sum_{i=1}^n \hat{\pi}^i \hat{r}^i \leq 0 \quad \text{and} \quad \sum_{i=1}^n \hat{\pi}^i \hat{c}^i \leq 0.$$

By definition of the Walrasian equilibrium and (b1),

$$(b3) \quad \rho \hat{r}^i + \gamma \hat{c}^i \geq \rho r^i + \gamma c^i, \quad \text{for all } i,$$

where the inequality is strict for some  $i$  (all  $i$  if  $\hat{\pi}$  is not strictly positive). Multiplying (b3) by  $\hat{\pi}^i$ , summing over  $i$  and recalling that the right-hand side of (b3) is zero for all  $i$ , we conclude that either  $\sum \hat{\pi}^i \hat{r}^i > 0$  or  $\sum \hat{\pi}^i \hat{c}^i > 0$ , a contradiction of (b2). Q.E.D.

**PROOF OF PROPOSITION 7:** Let us assume for convenience that  $n = 2$  (the argument extends to any number of types). Suppose contrary to the proposition that there exists an equilibrium with payoffs  $v = (v^1, v^2)$  that are not SUPO. For each vector of beliefs  $\hat{\pi}$  and for each corresponding Walrasian allocation  $\mu'_*(\hat{\pi})$ ,

$$(c1) \quad \text{there exists } i \text{ such that } \sum_j p_j V^i(\mu'_j(\hat{\pi})) > v^i.$$

<sup>26</sup> The reason for choosing  $r^1$  and  $c^1$  small is to ensure that the approximations (a1) and (a2) are good enough for (a3) to hold.

Moreover, generically, we have

$$(c2) \quad \sum_j p_j V^i(\mu'_j(\hat{\pi})) > \sum_j p_j V^i(\mu^h_j(\hat{\pi})) \quad \text{for all } i \neq h$$

and either

$$(c3) \quad U_2(\mu^1_2(\hat{\pi})) > \bar{u}$$

or

$$(c4) \quad U_2(\mu^1_2(\hat{\pi})) < \bar{u}.$$

Without loss of generality, assume that (c4) holds. Then

$$(c5) \quad U_2(\mu^2_2(\hat{\pi})) > \bar{u} \quad \left( \text{from (c4) and } \sum \hat{\pi}^i U_2(\mu^i_2(\hat{\pi})) = \bar{u} \right).$$

For each strictly positive vector  $\hat{\pi}$  and Walrasian allocation  $\mu^*(\hat{\pi})$ , choose  $\varepsilon > 0$  sufficiently small so that there exists a slight perturbation  $\hat{\mu}^*(\hat{\pi})$  such that (c1), (c2), (c4), (c5) remain true, and also

$$(c6) \quad \bar{v}^i < \sum_j p_j V^i(\hat{\mu}^i_j(\hat{\pi})) < \sum_j p_j V^i(\mu^i_j(\hat{\pi})) - \varepsilon \quad (i = 1, 2),$$

$$(c7) \quad \sum \hat{\pi}^i U_2(\hat{\mu}^i_2(\hat{\pi})) > \bar{u}, \quad \sum \hat{\pi}^i U_2(\hat{\mu}^i_2(\hat{\pi})) > \sum \hat{\pi}^i U_2(\hat{\mu}^i_1(\hat{\pi}))$$

and

$$(c8) \quad \sum \hat{\pi}^i U_1(\hat{\mu}^i_1(\hat{\pi})) > \sum \hat{\pi}^i U_1(\hat{\mu}^i_2(\hat{\pi})).$$

If  $\hat{\pi}^j = 1$ , then choose  $\varepsilon$  and  $\hat{\mu}^*(\hat{\pi})$  to satisfy the same conditions except (a) drop the left-hand inequality in (c6) for  $i = j$  and (b) for  $j = 2$ , impose  $U_1(\hat{\mu}^1_1(\hat{\pi})) = U_1(\hat{\mu}^2_2(\hat{\pi}))$  (instead of (c8)) and  $U_1(\hat{\mu}^1_1(\hat{\pi})) > U_1(\hat{\mu}^2_2(\hat{\pi}))$  (see the derivation of (c25) below for why we can impose this last inequality).

Condition (c6) ensures that  $\mu^*(\hat{\pi})$  Pareto-dominates  $\hat{\mu}^*(\hat{\pi})$  by at least  $\varepsilon$  (in utility terms). Conditions (c7) and (c8) require that the type 2 agent strictly prefers his perturbed Walrasian allocation to his reservation utility and that each type strictly prefers his own perturbed allocation to that of the other type. That a perturbation of  $\mu^*(\hat{\pi})$  can be found satisfying (c7) and (c8) is an immediate consequence of the sorting condition. That such a perturbation can also satisfy (c1) through (c6) (except (c3)) follows from continuity.

We must make sure that the mechanism we construct satisfies condition (b) of the definition of  $M^*$ . To this end, choose a countable dense subset  $\{v(1), v(2), \dots\}$  of the set of SUPO payoff vectors (such a selection would be unnecessary if we did not have to satisfy (b)). Because, generically, the SUPO allocations associated with given beliefs  $\hat{\pi}$  are locally unique, we can choose the subset  $\{v(1), v(2), \dots\}$  so that the corresponding beliefs  $\{\pi(1), \pi(2), \dots\}$  are all distinct. For  $t = 1, 2, \dots$ , let  $\hat{\mu}^*(\pi(t))$  be the Walrasian allocation associated with  $v(t)$  and let  $\hat{\mu}^*(\pi(t))$  be the corresponding perturbed Walrasian allocation satisfying (c1) through (c8) (except (c3)) above. Define

$$(c9) \quad \hat{\mu}^*(\hat{\pi}) = \begin{cases} \hat{\mu}^*(\hat{\pi}), & \text{if } \hat{\pi} = \pi(t), \\ \hat{\mu}^*(\hat{\pi}), & \text{otherwise, where } \hat{\mu}^*(\hat{\pi}) \text{ is a perturbation (satisfying} \\ & \text{(c1) - (c8) (except (c3)) of an arbitrary Walrasian} \\ & \text{allocation } \mu^*(\hat{\pi}). \end{cases}$$

For each  $\hat{\pi}$ , choose  $\hat{\mu}_3$  such that

$$(c10a) \quad U_1(\hat{\mu}_3^1) = U_1(\hat{\mu}_2^1),$$

$$(c10b) \quad U_1(\hat{\mu}_3^2) > U_1(\hat{\mu}_2^2),$$

$$(c11) \quad U_2(\hat{\mu}_2^i) > U_2(\hat{\mu}_3^i) \quad (i = 1, 2),$$

$$(c12) \quad \sum \hat{\pi}^i U_1(\hat{\mu}_1^i) > \sum \hat{\pi}^i U_1(\hat{\mu}_3^i),$$

and

$$(c13) \quad V^1(\hat{\mu}_3^1) - V^1(\hat{\mu}_3^2) > V^1(\hat{\mu}_1^1) - V^1(\hat{\mu}_1^2).$$

That we can satisfy (c10) through (c12) is a direct implication of continuity and the sorting condition. Formula (c13) follows if  $\mu_3^2$  entails a large enough monetary payment (which is simply thrown away) by the principal. There is no contradiction between (c10b) and (c13) because what the agent receives need not equal what the principal pays (we are slightly abusing notation by writing both the principal's and agent's utility as a function of the same allocation  $\hat{\mu}_3^2$ ). Similarly, take  $\hat{\mu}_4$  so that

$$(c14a) \quad U_2(\hat{\mu}_4^1) > U_2(\hat{\mu}_1^1),$$

$$(c14b) \quad U_2(\hat{\mu}_4^2) = U_2(\hat{\mu}_1^2),$$

$$(c15) \quad U_1(\hat{\mu}_1^1) > U_1(\hat{\mu}_4^1) \quad (i = 1, 2),$$

$$(c16) \quad \sum \hat{\pi}' U_2(\hat{\mu}_2^1) > \sum \hat{\pi}' U_2(\hat{\mu}_4^1),$$

and

$$(c17) \quad V^2(\hat{\mu}_4^2) - V^2(\hat{\mu}_4^1) > V^2(\hat{\mu}_2^2) - V^2(\hat{\mu}_2^1).$$

Finally, choose  $\hat{\mu}_5^1$  so that

$$(c18) \quad U_2(\hat{\mu}_5^1) > \bar{u} \quad (i = 1, 2),$$

$$(c19) \quad \sum_h \hat{\pi}^h U_j(\hat{\mu}_2^h) > U_j(\hat{\mu}_5^1) \quad (i, j = 1, 2),$$

$$(c20) \quad U_1(\hat{\mu}_2^1) > U_1(\hat{\mu}_5^1) \quad (i = 1, 2),$$

and

$$(c21) \quad V^2(\hat{\mu}_5^2) - V^2(\hat{\mu}_5^1) > V^2(\hat{\mu}_2^2) - V^2(\hat{\mu}_2^1).$$

To satisfy (c18) through (c20), we can choose  $\hat{\mu}_5^1$  so that  $y = 0$  and  $t$  is slightly positive (thus  $U_j(\hat{\mu}_5^1) \approx \bar{u}$ ,  $i, j = 1, 2$ ). To ensure (c21) we can, as above, require a large monetary payment by the principal should she set  $\hat{\alpha} = \alpha^1$ .

Consider the following contract  $m^*$ . In this contract, a third party first announces a vector of "beliefs,"  $\hat{\pi}$ , which, in equilibrium, will turn out to be (at least approximately) the agent's beliefs about the principal's type. The principal and the agent then make simultaneous announcements about their types. That is, the principal announces  $\hat{\alpha}' \in \{\alpha^1, \alpha^2\}$ . Because we have added three more "types" for the agent—corresponding to  $\hat{\mu}_3^1$ ,  $\hat{\mu}_4^1$ , and  $\hat{\mu}_5^1$ —the agent's announcement  $\hat{\theta}$  lies in  $\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$ . If the announcements are  $\hat{\pi}$ ,  $\hat{\alpha} = \alpha'$ , and  $\hat{\theta} = \theta_j$ , the contract specifies the allocation  $\hat{\mu}_j'(\hat{\pi})$  (defined by (c9)). Moreover, if  $j \in \{1, 2\}$ , the third party is given a (small) monetary payoff (such a payoff is feasible since  $\hat{\mu}_j'(\hat{\pi})$  does not quite attain the Walrasian allocation for  $\hat{\pi}$ ) and nothing if  $j \in \{3, 4, 5\}$ .

We shall argue that, if the principal proposes  $m^*$ , the agent will accept it. There exists an equilibrium of  $m^*$  in which the third party announces  $\hat{\pi}$  equal to the agent's true beliefs  $\hat{\pi}$  and in which the principal and agent both announce their types truthfully. Moreover, in *any* equilibrium of  $m^*$  the principal and agent are truthful. Therefore, the only possible allocations resulting from proposing  $m^*$  are the  $\hat{\mu}_j'(\hat{\pi})$ . But (c1) implies that, for any  $\hat{\pi}$ , there exists at least one type  $i$  of principal who prefers  $\hat{\mu}_i'(\hat{\pi})$  to  $v^1$ . Thus non-SUPO equilibrium allocations are impossible.

We first demonstrate that  $\hat{\pi} = \hat{\pi}$  and truthtelling by the principal and agent constitute an equilibrium. Notice that, if  $\hat{\pi} = \hat{\pi}$  and the principal is truthful, (c7), (c8), (c11), (c12), (c15), (c16), (c19), and (c20) imply that the agent is truthful. Moreover, (c2) guarantees that the principal is truthful if the agent is. Now, if the third party announces  $\hat{\pi} = \hat{\pi}$  and the principal and agent are truthful, the third party's payoff is maximal, since the probability that the agent announces  $\theta_1$  or  $\theta_2$  is one. Hence, there is no other announcement he could make that could possibly raise his payoff.

We next show that, if  $\hat{\pi} = \hat{\pi}$ , the *only* possible equilibrium is the truthful one. Suppose, to the contrary, that there is an untruthful equilibrium. In this equilibrium, let  $\pi_{\star}^i$ ,  $i = 1, 2$ , be the probability that  $\hat{\alpha} = \alpha^i$ . Now, if  $(\pi_{\star}^1, \pi_{\star}^2) = (\hat{\pi}^1, \hat{\pi}^2)$ , then the argument from the preceding paragraph implies that the agent is truthful, which in turn implies that the principal is truthful, a contradiction. Hence  $(\pi_{\star}^1, \pi_{\star}^2) \neq (\hat{\pi}^1, \hat{\pi}^2)$ .

CASE I:  $\pi_{\star}^2 > \hat{\pi}^2$  and  $\pi_{\star}^1 < \hat{\pi}^1$ .

Note first that (c11) implies that the type 2 agent does not set  $\hat{\theta} = \theta_3$ . Now, if  $V^2(\hat{\mu}_2^2) \geq V^2(\hat{\mu}_1^2)$ , then in view of (c5), the allocation  $\hat{\mu}_2^2$  satisfies the agent's IR and (trivially) IC constraints and yet, from (c6), generates more utility for the type 2 principal than  $\bar{v}^2$ , an impossibility.

Hence,

$$(c22) \quad V^2(\hat{\mu}_1^2) > V^2(\hat{\mu}_2^2).$$

If  $\Sigma \pi_{\star}' U_2(\hat{\mu}_1^1) \geq \Sigma \pi_{\star}' U_2(\hat{\mu}_2^1)$ , then, in view of  $\pi_{\star}^2 > \hat{\pi}^2$  and the second inequality of (c7),  $U_2(\hat{\mu}_1^2) \geq U_2(\hat{\mu}_2^2)$ . But if this last inequality held, then, because of (c22),  $\hat{\mu}_1^2$  would satisfy the agent's IR and (trivially) IC constraints but generate a higher payoff than  $\bar{v}^2$  for the type 2 principal, an impossibility. We conclude that

$$(c23) \quad U_2(\hat{\mu}_1^2) < U_2(\hat{\mu}_2^2),$$

$\Sigma \pi_{\star}' U_2(\hat{\mu}_1^1) < \Sigma \pi_{\star}' U_2(\hat{\mu}_2^1)$ , and therefore, that the type 2 agent does not set  $\hat{\theta} = \theta_1$ . In turn, (c23) implies, together with (c14b) and (c16), that the type 2 agent does not set  $\hat{\theta} = \theta_4$ . From (c11), he does not take  $\hat{\theta} = \theta_3$ . Finally, (c4) and (c5) imply that  $U_2(\hat{\mu}_2^2) > U_2(\hat{\mu}_1^2)$ , and so  $\pi_{\star}^2 > \hat{\pi}^2$  and (c19) imply that the type 2 agent does not set  $\hat{\theta} = \theta_5$ . We conclude that the type 2 agent sets  $\hat{\theta} = \theta_2$ .

From (c10a) and (c10b) and the fact  $\pi_{\star}^1 < \hat{\pi}^1$ , the type 1 agent does not set  $\hat{\theta} = \theta_2$ . From (c15) he does not announce  $\hat{\theta} = \theta_4$ . Finally, (c20) implies that he does not set  $\hat{\theta} = \theta_5$ . Hence, the type 1 agent must either set  $\hat{\theta} = \theta_1$  or  $\hat{\theta} = \theta_3$  (or randomize between them). In the former case, given that the type 2 agent is truthful, the type 1 principal is clearly better off announcing  $\hat{\alpha} = \alpha^1$ . But (c13) implies that she also is better off with  $\hat{\alpha} = \alpha^1$  in the latter case (and hence for any randomization between  $\theta_1$  and  $\theta_3$ ). Thus, in equilibrium, the type 1 principal takes  $\hat{\alpha} = \alpha^1$ , contradicting  $\pi_{\star}^1 < \hat{\pi}^1$ .

CASE II:  $\pi_{\star}^2 < \hat{\pi}^2$  and  $\pi_{\star}^1 > \hat{\pi}^1$ .

Note first that, if  $\hat{\pi}^2 < 1$ ,

$$(c24) \quad U_1(\hat{\mu}_1^2) < U_1(\hat{\mu}_2^2);$$

otherwise, in view of (c5), the allocation  $\hat{\mu}_2^2$  satisfies the agent's IC and IR constraints. But  $\hat{\mu}_2^2$  gives the type 2 principal more utility than  $\bar{v}^2$ , a contradiction. Now, (c24) and (c8) imply that

$$(c25) \quad U_1(\hat{\mu}_1^1) > U_1(\hat{\mu}_2^1)$$

for  $\hat{\pi}^2 < 1$ . Now, from our choice of  $\hat{\mu}_{\cdot}(\hat{\pi})$  (see the passage following (c8)), (c24) becomes an equation and (c25) is satisfied even when  $\hat{\pi}^2 = 1$ . In view of (c8), (c24) (both when  $\hat{\pi}^2 < 1$  and  $\hat{\pi}^2 = 1$ ), and (c25),  $\pi_{\star}^1 > \hat{\pi}^1$  implies that

$$(c26) \quad \Sigma \pi_{\star}' U_1(\hat{\mu}_1^1) > \Sigma \pi_{\star}' U_1(\hat{\mu}_2^1),$$

and so the type 1 agent will not set  $\hat{\theta} = \theta_2$ . This, in turn, together with (c10a), (c10b), and (c12), implies that he will not announce  $\hat{\theta} = \theta_3$ . Now, (c15) implies that the type 1 agent will not set  $\hat{\theta} = \theta_4$ . Finally (c20) implies that he will not announce  $\hat{\theta} = \theta_5$ . We conclude that

$$(c27) \quad \text{the type 1 agent must announce } \hat{\theta} = \theta_1.$$

From (c14a) and (c14b) and the fact that  $\pi_{\star}^2 < 1$ , the type 2 agent will not announce  $\hat{\theta} = \theta_1$ , and from (c11) he will not announce  $\hat{\theta} = \theta_3$ . Hence, his announcement must be  $\theta_2$ ,  $\theta_4$ , or  $\theta_5$  (or some randomization among them). Now, if he announces  $\theta_2$ , then, from (c27) and (c2), it is optimal for the type 2 principal to announce  $\hat{\alpha} = \alpha^2$ . But from (c17) and (c21) the same is true for announcement  $\theta_4$  or  $\theta_5$ . Hence,  $\hat{\alpha} = \alpha^2$  must be the type 2 principal's equilibrium announcement, contradicting  $\pi_{\star}^2 < \hat{\pi}^2$ . We conclude that, if  $\hat{\pi} = \hat{\pi}$ , only a truthful equilibrium is possible.

We next observe that in any equilibrium where the agent never makes an announcement other than  $\theta_1$  or  $\theta_2$ , both the principal and agent are truthful. To see this, note that if in equilibrium the type 1 agent announces  $\hat{\theta} \in \{\theta_1, \theta_2\}$ , then (c10a) and (c10b) rule out the choice  $\hat{\theta} = \theta_2$  unless  $\hat{\pi}^1 = 1$ , in which case (c25) rules out this choice. Moreover, if the type 2 agent announces  $\hat{\theta} \in \{\theta_1, \theta_2\}$ , then (c14a) and (c14b) imply that  $\hat{\theta} = \theta_2$ , unless  $\hat{\pi}^2 = 1$ , in which case (c23) ensures  $\hat{\theta} = \theta_2$ . Hence, the agent is truthful in equilibrium, which in turn implies that the principal must be.

Now, the third party maximizes his payoff by maximizing the probability that the agent announces  $\hat{\theta} \in \{\theta_1, \theta_2\}$ . But, as we have seen, the party can ensure that the probability is 1 by setting  $\hat{\pi} = \hat{\pi}$ . This

does not imply that the third party *does* set  $\hat{\pi} = \hat{\pi}$  in equilibrium because the truthtelling constraints continue to hold for  $\hat{\pi}$  somewhat different from  $\hat{\pi}$ . But it does mean that  $\hat{\pi}$  is approximately  $\hat{\pi}$ . In any case, the only possible equilibria are those where the agent never makes an announcement other than  $\theta_1$  or  $\theta_2$ . But, as we have shown, this implies that the principal and agent are truthful. Hence, any equilibrium allocation is in the class  $\hat{\mu}(\hat{\pi})$ . Note finally that (c18) implies that if the principal proposes  $m^*$ , the agent will accept it.

The fact that there always exists an equilibrium of  $m^*$  (where  $\hat{\pi} = \hat{\pi}$  and the principal and agent are truthful) implies that  $m^*$  satisfies condition (a) of the set of admissible mechanisms  $M^*$ . To see that it satisfies condition (b), consider SUPO payoffs  $(\hat{v}_1, \hat{v}_2)$ . If  $(\hat{v}_1, \hat{v}_2) = v(t)$  for some  $t$ , then (c6) implies that for beliefs  $\pi(t)$ ,  $(\hat{v}^1, \hat{v}^2)$  Pareto-dominates the truthful equilibrium payoffs for  $m^*$ . Moreover as long as, for some  $t$ ,  $|\hat{v}^i - v^i(t)| < (\epsilon/2)$  for all  $i$ , (c6) implies the same conclusion. Now, this last inequality is satisfied because the  $v(t)$ 's are dense in the set of SUPO payoffs. Hence, condition (b) is satisfied, and  $m^*$  belongs to  $M^*$ . Q.E.D.

The arguments in the proof of Proposition 7 are somewhat involved, but the mechanism  $m^*$  is quite simple. The third party first announces the agent's beliefs about the principal's type and then the principal and agent announce their types. The allocations  $\mu_3^*$ ,  $\mu_4^*$ , and  $\mu_5^*$  simply ensure that, in equilibrium, the principal and agent announce their types truthfully if the third party announces the agent's true beliefs.

**PROOF OF PROPOSITION 10:** That a Walrasian allocation relative to  $\pi$  is an FGP equilibrium is trivial and results from the fact that Walrasian equilibria belong to the core. To prove the converse, let  $v^*$  denote an FGP equilibrium payoff vector and consider the correspondence from the set of beliefs  $\hat{\pi}$  and feasible payoffs  $v$  into itself:

$$(\hat{\pi}, v) \rightarrow \left\{ \begin{array}{l} \hat{\pi} | \hat{\pi}'(v' - v^*)k \geq 0 \text{ for all } i, \\ \frac{\hat{\pi}^i}{\pi^i} \max[0, v^j - v^{j*}]k \leq \frac{\hat{\pi}^j}{\pi^j} \max[0, v^j - v^{j*}]k \text{ for all } i \text{ and } j, \\ \text{where } k = \max \left[ 0, \max_h (v^h - v^{h*}) \right] \end{array} \right\} \\ \times \{ \tilde{v} | \tilde{v} \text{ is the Walrasian payoff vector relative to } \hat{\pi} \}.$$

By construction, a fixed point of this nonempty, upper hemicontinuous and convex-valued correspondence puts zero weight on types who "lose" relative to the equilibrium ( $v' < v^{j*}$ ), and preserves relative weights with respect to prior beliefs (i.e.,  $(\hat{\pi}^i/\pi^i) = (\hat{\pi}^j/\pi^j)$  for types who are strictly better off ( $v' > v^{j*}$  and  $v^j > v^{j*}$ )). (Note also that  $\hat{\pi}^i/\pi^i \leq \hat{\pi}^j/\pi^j$  if  $v^i = v^{i*}$  and  $v^j > v^{j*}$ .) If, for this fixed point,  $v' \leq v^{i*}$  for all  $i$ , then the equilibrium payoff vector is Walrasian. If there exists  $i$  such that  $v^i > v^{i*}$ , then by construction  $k > 0$  and the equilibrium payoff vector  $v^*$  does not satisfy the FGP refinement. Q.E.D.

## REFERENCES

- ADMATI, A., AND M. PERRY (1986): "Strategic Delay in Bargaining," mimeo.  
 AUMANN, R. (1974): "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1, 67–96.  
 BARON, D., AND R. MYERSON (1982): "Regulating a Monopolist with Unknown Costs," *Econometrica*, 50, 911–930.  
 CRAWFORD, V. (1985): "Efficient and Durable Decision Rules: A Reformulation," *Econometrica*, 53, 817–837.  
 DASGUPTA, P., P. HAMMOND, AND E. MASKIN (1979): "The Implementation of Social Choice Rules," *Review of Economic Studies*, 46, 185–216.  
 DEBREU, G. (1959): *The Theory of Value*. New Haven: Yale University Press.  
 FARRELL, J. (1985): "Credible Neologisms in Games of Communication," mimeo, MIT.

- FUDENBERG, D., AND J. TIROLE (1983): "Sequential Bargaining with Incomplete Information," *Review of Economic Studies*, 50, 221–247.
- (1989): "Perfect Bayesian and Sequential Equilibria," mimeo, MIT.
- GREEN, J., AND J.-J. LAFFONT (1979): *Incentives and Public Decision-Making*. Amsterdam: North-Holland.
- GROSSMAN, S., AND O. HART (1981): "Implicit Contracts, Moral Hazard, and Unemployment," *American Economic Review Papers and Proceedings*, 71, 301–307.
- GROSSMAN, S., AND M. PERRY (1986): "Perfect Sequential Equilibrium," *Journal of Economic Theory*, 39, 97–119.
- GUESNERIE, R., AND J.-J. LAFFONT (1984): "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm," *Journal of Public Economics*, 25, 329–369.
- HIRSHLEIFER, J. (1971): "The Private and Social Value of Information and the Reward to Inventive Activity," *American Economic Review*, 61, 561–574.
- KREPS, D., AND R. WILSON (1982): "Sequential Equilibrium," *Econometrica*, 50, 863–894.
- LAFFONT, J.-J., AND E. MASKIN (1982): "The Theory of Incentives: An Overview," in *Advances in Economic Theory*, ed. by W. Hildenbrand. Cambridge: Cambridge University Press, 31–94.
- LAFFONT, J.-J., AND J. TIROLE (1986): "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 94, 614–641.
- MAS-COLELL, A. (1985): *The Theory of General Economic Equilibrium: A Differential Approach*. New York: Cambridge University Press.
- MASKIN, E., AND J. RILEY (1984): "Monopoly with Incomplete Information." *Rand Journal of Economics*, 15, 171–196.
- MASKIN, E., AND J. TIROLE (1986): "Principals with Private Information, I: Independent Values," Harvard University Discussion Paper #1234.
- (1988): "The Principal-Agent Relationship with an Informed Principal, II: Common Values," mimeo, M.I.T.
- MIRRELES, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38, 175–208.
- MUSSA, M., AND S. ROSEN (1979): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–317.
- MYERSON, R. (1979): "Incentive Compatibility and the Bargaining Problem." *Econometrica*, 47, 61–73.
- (1983): "Mechanism Design by an Informed Principal." *Econometrica*, 51, 1767–1798.
- SOBEL, J., AND I. TAKAHASHI (1983): "A Multi-Stage Model of Bargaining," *Review of Economic Studies*, 50, 411–426.
- SPENCE, A. M. (1974): *Market Signaling*. Cambridge: Harvard University Press.