Discussion Paper No. 850

RENEGOTIATION-PROOF IMPLEMENTATION AND TIME PREFERENCES*

by

Ariel Rubinstein**
Department of Economics
The Hebrew University of Jerusalem

and

Asher Wolinsky
Department of Economics
Northwestern University

April 1989

Revised August 1989

---

*This is an early draft; we shall be grateful for comments. We thank D. Abreu, A. Ma, S. Matthews, J. Moore and D. Pearce for useful comments.

**This author was visiting the Department of Economics at the University of Chicago and the London School of Economics during the period in which the research for this paper was conducted.

Renegotiation-Proof Implementation and Time Preferences

Abstract

This paper explores how the requirement that the implementation of contracts be renegotiation proof affects the set of contracts which theoretically can be implemented in a seller-buyer scenario in which the information regarding the agents' valuations is non-verifiable. The paper's main contribution is that, first, it explicitly adds a time dimension to an implementation problem, and second, it introduces a natural criterion of renegotiation proofness for the case of time consuming renegotiation. Consequently, the results regarding the set of implementable contracts are different from those in the related literature.

Renegotiation-Proof Implementation and Time Preferences

1. Introduction

This paper explores how the requirement that the implementation of contracts be renegotiation proof affects the set of contracts which theoretically can be implemented in a seller-buyer scenario in which the information regarding the agents' valuations is non-verifiable. The paper's main contribution is that, first, it explicitly adds a time dimension to an implementation problem, and second, it introduces a natural criterion of renegotiation proofness for the case of time consuming renegotiation. Consequently, the results regarding the set of implementable contracts are different from those in the related literature.

For concreteness, the discussion will be conducted in the context of the following example. There are two agents, a seller and a buyer, who sign a contract for the sale of one unit in the future. Their valuations for the unit, denoted by s and b respectively, are not known when they sign the contract. They become known to both parties after the contract is signed and before it is implemented. Thus, when it comes to implementing the contract the parties' information is complete, but it is assumed to be non-verifiable, i.e., not observable to third parties (such as a court). For the ease of exposition, we first restrict attention to the case in which in all realizations of s and b, $b>s$ (this restriction will be dropped in Appendix 1). When they sign the contract, the parties want to specify, for any possible realization of the valuations (s,b), whether or not there will be trade and the price $P(s,b)$ at which it will take place. The contract describes the procedure that will be followed after the valuations are realized. The purpose of this procedure is to make sure that the original intentions of the parties are indeed carried out. The idea is that the steps laid out in the contract are independent of (s,b) and hence are enforceable by a third party (or by a social convention). In game theoretic terms, a contract specifies a game form that implements the function P in the sense that, for all s and b, P(s,b) is the resulting game's unique subgame perfect equilibrium outcome.

In the context of this scenario we shall investigate the set of price functions P(s,b) which can be so implemented by a contract and examine how this set is affected by the requirement that the contract be renegotiation-proof. The senses in which we use the term renegotiation proof will be made precise below, but roughly speaking it means that the contract is such that in no stage the parties find it mutually beneficial to scrap it and reach an alternative agreement.

As already mentioned above, we are mainly interested in the idea of introducing explicitly the time dimension into an implementation problem and using it to look at renegotiation-proof implementation. The contracting scenario outlined above is therefore used mainly as a vehicle to expose these ideas, and so we devote relatively less space to discussing the economic issues involved. Nevertheless, before we proceed, let us point out briefly why the scenario and the questions to be considered are interesting from the viewpoint of economic analysis. Suppose that the seller's valuation is zero, that the buyer's valuation can be either 1 or 2 and that ex-ante there is equal probability for anyone of the buyer types. Assume that in order to produce the unit the seller has to invest 1.3, before the buyer's valuation is determined. Notice that a contract that specifies the constant price of

1.4, regardless of the buyer's valuation, achieves the efficient outcome that the seller produces the unit and that it ends up at the hands of the buyer. However, this contract is not ex post individually rational and the low valuation buyer will block the sale if he has the power to do so. Thus, in an environment where outcomes that are not ex post individually rational cannot be enforced, a constant price contract will not induce the seller to make the necessary initial investment and the overall outcome will be inefficient. The question of whether this situation necessarily gives rise to inefficient under-investment amounts, therefore, to inquiring whether or not it is possible to implement other price functions, such as the one that prescribes prices 0.8 and 1.8 to the low and high valuation buyers respectively. And if the situation is such that the parties are free to renegotiate, the relevant question is whether such other price functions can be implemented by a renegotiation-proof contract. Of course, by identifying the set of all price functions implementable by a renegotiation-proof contract, we get at once the answer to this question and to similar ones. In this sense, the problems investigated in the present paper, regarding the entire sets of price functions which are so implementable, are potentially useful for the analysis of a class of cases of this type.

In the first case considered the possible outcomes which may be reached in the implementation game are sale at a certain price or the "no-sale" outcome. Our initial result (proposition 0) establishes that it is possible to implement a big set of functions, including all functions P which are increasing[1] in s and b and satisfying $b>P(s,b)>s$. However, the mechanism constructed in the proof of proposition 0 makes use of the inefficient "no-sale" outcome, i.e., certain out of equilibrium moves in the implementation game will lead to "no sale". This feature is questionable when one thinks of a voluntary contract, since in situations in which agents are sovereign to mutually agree to scrap the mechanism, they will probably not put up with substantial inefficiency and instead will negotiate a new outcome. Therefore, if agents see through the contract and anticipate renegotiation, they will not be necessarily deterred by the no-sale outcome and hence the contract might not achieve the desired outcomes.

This criticism motivates the work of Maskin and Moore(1988), Green and Laffont(1988) and Aghion, Dewatripont and Rey (1989). They respond to it by looking at contracts which take into account the negotiated outcome that will follow an inefficient one. We follow here a somewhat different approach and look at contracts which are immune to this criticism. Such contracts are called renegotiation-proof. The first notion of renegotiation-proofness that we discuss requires that, for all s and b and in all subgames of the implementation game, the SPE outcomes be efficient. In the environment considered here, this requirement amounts to ruling out the no-sale outcome as a possible SPE outcome of the implementation game, whenever $b>s$. The idea is that if the game were to end with this outcome it could not be considered renegotiation-proof since this outcome would be renegotiated to an agreement which is preferred by both. Proposition 1 shows that this renegotiation proofness criterion indeed restricts considerably the set of implementable contracts to include only those that specify a constant price irrespective of (s,b).

The fact that renegotiation-proof contracts form a limited subset of the set of possible contracts has been presented in the literature as a source of inefficiency, since this limited subset may not contain sufficiently rich contracts which are required under certain circumstances to provide the right incentives for, say, investment which has to take place before the information is revealed (see, e.g., Green and Laffont(1988) Hart and Moore(1988)).

The substantial limitation of the set of admissible contracts described above seems to be an artifact of the too stringent criterion of renegotiation proofness employed. The idea of eliminating all inefficient outcomes implicitly assumes that there is time dimension and that, after the implementation game is over, the parties turn to renegotiate inefficient outcomes. But the above approach leaves this dimension unmodelled and does not specify what the time structure is and how the fact that time is normally costly figures into the considerations.

We modify the model by adding explicitly the time dimension. The set of possible outcomes will now be richer since outcomes will be dated so that a typical outcome is a pair (p,t) with the interpretation that the good is sold for price p at time t. The mechanism described by the contract should be interpreted as including a time table of the different steps in the execution of the contract. The two features of the time dimension which are relevant for the problem at hand are, first, that delays are costly and, second, that these costs are irretrievable--it is impossible to go back in time.

The notion of renegotiation-proofness to be used here requires that in the beginning of each period, after any possible history, the subgame perfect equilibrium outcome is Pareto efficient. That is, the parties will not find it mutually advantageous to renegotiate another outcome at the beginning of any period. In other words, if one period is the minimum amount of time required to renegotiate a contract, then at no point will both parties want to renegotiate.

It turns out that with this notion of renegotiation proofness, all contracts that specify for $b>s$ trade at the price $P(s,b)$, where $P(s,b)$ is increasing in s and b and $s<P(s,b)<b$, are implementable in a renegotiation proof manner much as they are in the absence of renegotiation proofness requirements, and in contrast to the case in which only efficient outcomes are considered renegotiation proof. That is, modeling renegotiation proofness by restricting attention only to efficient outcomes is by no means an approximation to what goes on in an environment with costly renegotiation, even if the cost is small.

Thus, since results in the spirit of proposition 1 can lead to conclusions on inefficient behavior, the last result suggests that such explanations may not be valid when the contract can use the time dimension and when recontracting is costly (time consuming).

Finally, we remark that modeling implementation over time requires looking into implementation by extensive game forms, where the natural solution concept is SPE. The study of implementation by SPE was started by Moore and Repullo(1988).

2. <u>The Model</u>

There are a seller and a buyer who are interested in signing a contract for the sale of a certain unit. The reservation values, denoted by s and b respectively, are taken from the sets S and B. We assume that S and B have both maximum and minimum. Let s_{max}, b_{max}, s_{min} and b_{min} denote the maxima and minima of S and B respectively. For the ease of exposition, throughout the paper we assume that $b_{min}>s_{max}$ so that there is always room for trade. In Appendix I we shall extend the definitions and the results to the case in which S and B may overlap so that we may have $s_{max}>b_{min}$. When the contract is signed the parties know only S and B, but before it is carried out the true values of s and b are realized and are common knowledge between both parties.

The possible outcomes with which this interaction may end are exchanges for some price p, and the no-sale outcome. We shall refer to an exchange at price p as outcome p and to the no-sale outcome as outcome D (for "disagreement"). Notice that this specification of possible outcomes restricts the range of possible contracts, e.g, it does not include contracts such that, after certain developments, one party or both pay penalties to or receive subsidies from a third party. The preferences of the parties over these outcomes are given by the utility functions p-s and b-p for the seller and the buyer respectively; utility of zero is assigned by both to the no sale outcome.

We shall be interested in implementing by a contract a price function P that assigns a price P(s,b) to each pair s,b. We shall further restrict the discussion only to price functions which are strictly ex post individually rational, i.e., satisfy $s<P(s,b)<b$.

Definition 1: A price function P(s,b) will be called implementable if there exists a game form with perfect information such that: (1) all its terminal nodes are either (exchanges for) prices in the interval $[s_{min},b_{max}]$, or the no-sale outcome D; (2) for all s and b, the unique subgame perfect equilibrium outcome is an exchange at the price P(s,b).

The interpretation of the game form is that of a procedure fixed by a contract which is signed before s and b are realized and can be carried out after these values are realized, if one of the parties wants. Since when it comes to implementing the contract both parties know s and b, the implementation game is one of complete information. However, the meaning of the requirement that one game form implements P(s,b) for all s and b is that s and b are unobservable to third parties, such as a court that enforces the steps prescribed by the contract. (Alternatively, if the game is a social custom rather than a formal procedure implemented by court, this requirement means that the custom applies to all cases independently of the reservation values.)

Notice that we adopt here a specific notion of implementability out of a number of such possible concepts. Possible variations on the definition of implementability would either relax the uniqueness requirement or replace the SPE with another solution concept.

3. A Preliminary Result: Implementation Without Renegotiation Proofness

The first result prepares the background for our later discussion in renegotiation proofness. It demonstrates that the set of functions P(s,b) that are implementable is rich. The ideas of the proof are related to those presented in the literature on Subgame Perfect Implementation by Moore and Repullo (1988) and Glazer and Ma (1988).

Proposition 0: Any function P, $s<P(s,b)<b$, which is increasing in s and b is implementable.

Proof: Let P be a function of s and b, $s<P(s,b)<b$, increasing in both arguments. Consider the following game in extensive form.

Stage 1: The announcement stage

The seller announces a number v_S in S.
The buyer announces a number v_B in B or challenges the seller.
If the buyer challenges the seller, the game continues to stage 2.
If the buyer chose v_B, the seller may challenge the buyer: if the seller does, the game continues to stage 3; if he does not, the unit is exchanged for $P(v_S,v_B)$.

Stage 2: The buyer can make a "take it or leave it" price offer below v_S

The buyer can choose a price offer $p<v_S$.
The seller either accepts in which case the good is exchanged for p, or rejects in which case the outcome is D.

Stage 3: The seller can make a "take it or leave it" price offer above v_B

The seller can choose a price offer $p>v_B$.
The buyer either accepts and then the good is exchanged for p, or rejects and then the outcome is D.

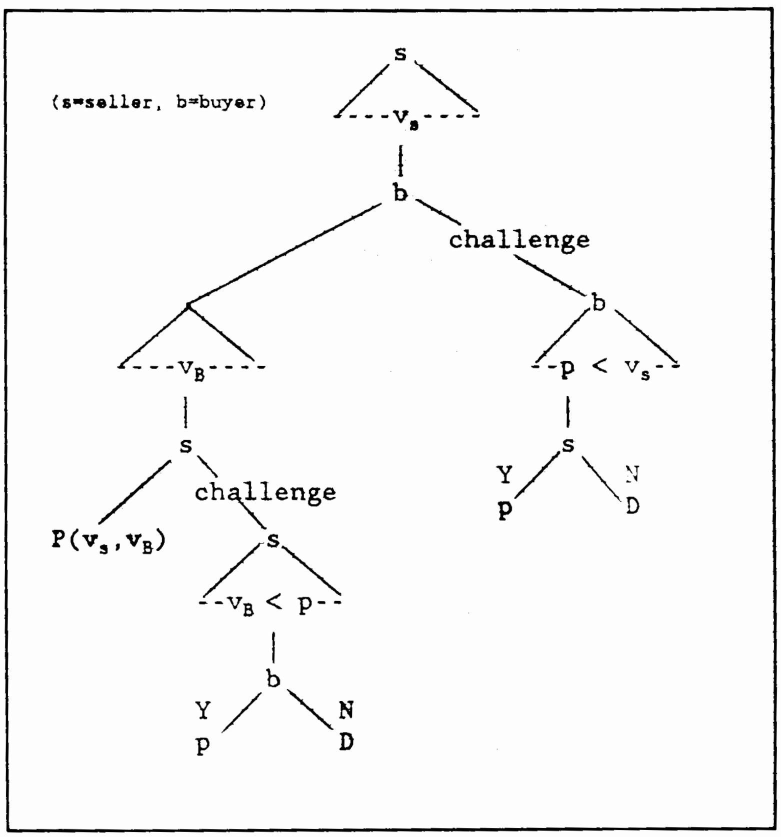

Figure 1

Let us verify that this game form indeed implements P.

Step 1: Consider a subgame starting at stage 2. If $v_S>s$ the SPE outcome is exchange for the price s and if $v_S \le s$, the SPE outcome is D. Similarly, for a subgame that starts at stage 3, if $v_B<b$ the SPE outcome is exchange for the price b and if $v_B \ge b$, the SPE is D.

Step 2: In the subgame that starts after the seller announced $v_S>s$, there are the following three possibilities. If the buyer challenges, then by step 1 the SPE outcome is an exchange at the price s; if he does not challenge and announces $v_B<b$, by step 1 the outcome is an exchange at b; and if he announces $v_B \ge b$, the outcome is $P(v_S,v_B) \ge P(v_S,b) \ge P(s,b) > s$. Thus, in any SPE the buyer challenges and the outcome is an exchange at s.

Step 3: In the subgame that starts after the seller announced $v_S<s$, the buyer can announce $v_B=b$ after which the SPE outcome is exchange at the price $P(s,b) \le P(v_S,b)$.

Step 4: Consider the different continuations in the subgame that starts after the seller announced $v_S=s$. If the buyer challenges, the SPE will be D. If the buyer announces $v_B=b$, the seller will not challenge him in the following SPE and the outcome will be an exchange at the price P(s,b). If the buyer announces $v_B<b$ the SPE outcome will be exchange for the price b which is worse for the buyer than the price P(s,b). Finally, if the buyer announces $v_B>b$, the SPE outcome in the continuation will be $P(s,v_B) \ge P(s,b)$. Thus, after the seller announces $v_S=s$, the SPE outcome will be exchange for P(s,b).

From steps 1,2,3 and 4 above it follows that the only SPE outcome is exchange for P(s,b).

QED

Thus, given that the set of possible outcomes of the implementation game consists only of the exchanges and the no-sale outcome, all price functions P, $s<P(s,b)<b$, which are increasing in s and b are implementable by a contract of the type considered here. Looked upon from the point of view of contract design, this is an optimistic result since a wide range of price functions can be implemented.

Although we shall not be concerned here with characterizing the exact set of implementable functions, let us point out that this condition on the monotonicity of P is "almost" a necessary condition in the following sense. Any function P which is implementable by a finite game form (i.e., the number of nodes in the game tree is finite) is increasing in s and b. We show this in Appendix II by proving a lemma stating that if P is implementable and $P(s,b)<P(s,b')$ for some $b>b'$, then the implementation game between s and b' must have another SPE with outcome D or price not higher than P(s,b), in contradiction to the implementability of P (which by definition requires uniqueness of the SPE outcome).

4. Renegotiation Proofness is Identified with Efficiency

The proof of Proposition 0 relies on the possibility to enforce the no-sale outcome, D: if the seller lies about the buyer's valuation, then the equilibrium outcome in the resulting subgame is D. Notice that since $b>s$ outcome D is inefficient. Thus, if the situation is such that the parties can communicate and are sovereign to scrap the old contract, then in the event that D is indeed reached they would probably renegotiate a mutually beneficial exchange. The implied criticism is that proposition 0 might exaggerate the set of implementable price functions. If a contract involves inefficient outcomes, the parties will presumably see through it and base their decisions on the anticipated outcomes of the renegotiation. Therefore, the parties will not necessarily be prevented from taking steps that lead to outcome D and the contract may fail to implement some price functions.

The above argument suggests that in order to investigate what price functions can really be implemented we should look at contracts which are immune to criticism of this type, i.e., renegotiation-proof contracts. The following definition gives the first notion of renegotiation-proofness considered here.

Definition 2: A price function P(s,b) will be called renegotiation-proof implementable if: (1) it is implementable; (2) the no-sale outcome cannot be reached after a finite number of moves; (3) for all s and b and all subgames, the SPE outcome is efficient.

This definition essentially requires that at no node of the implementation game will it be mutually beneficial for the parties to renegotiate. As we have already mentioned Green and Laffont(1988), Maskin and Moore(1986) and Aghion, Dewatripont and Rey(1989) take a somewhat different approach to this problem. They do not require the contracts themselves to be renegotiation-proof as we do here, but rather study their consequences in the presence of exogenous renegotiation technologies (for more details see the part of the discussion in section 6).

The following proposition shows how this requirement of renegotiation proofness reduces considerably the set of implementable price functions and hence the range of possible contracts.

Proposition 1:

The only renegotiation proof implementable (strictly ex post individually rational) price functions are the constant functions $P(s,b)=p$, where $s_{max}<p<b_{min}$.

Proof:

Suppose that the function $P(s,b)$ such that $s<P(s,b)<b$ is renegotiation-proof implementable. Observe that, when outcome D is ruled out, the preferences of all buyers over the remaining outcomes coincide and likewise the preferences of all sellers coincide as well. Therefore, any equilibrium in the game between s and b is also an equilibrium in the game between s′ and b′, and if an exchange at $P(s,b)$ is the unique SPE outcome in the game between s and b, it must be the unique SPE in the game between s′ and b′. It follows that $P(s,b)=p$ where the constant p is in (s_{max},b_{min}).

QED

Note that while the reason for ruling out inefficient outcomes from being SPE in any subgame is that they will be renegotiated, the approach described in this section does not model explicitly the renegotiation process. Instead it implicitly assumes that renegotiation is costless and is always concluded successfully. The question is how sensitive the result is to this abstraction--whether, for example, the above result changes significantly once we recognize that renegotiation could be costly.

5. Renegotiation and Time

The model considered in the two previous sections is rather crude: the basic outcomes are either efficient (an immediate agreement) or grossly inefficient (no-sale). We shall consider now a more refined model of the situation and the added detail will result in a richer set of outcomes.

Specifically, we assume that the model has a time dimension. Time is divided into discrete periods 0,1,2..., where period 0 corresponds to the point at which the implementation game begins. The set of possible outcomes includes all outcomes of the form: "the good is sold for price p at time t", to be denoted (p,t), and the no-sale outcome, D. The parties' preferences $\gtrsim_s$, $\gtrsim_b$ extend the preferences over the basic outcomes and are assumed to have the following three properties: (i) impatience: for all $p>s$, $(p,t)\gtrsim_s(p,t+1)$, and for all $p<b$, $(p,t)\gtrsim_b(p,t+1)$; (ii) substitution: for any $p>s$ and t, there exists $q<p$ such that $(q,t)\gtrsim_s(p,t+1)$, and for any $p<b$ and t, there exists $q>p$ such that $(q,t)\gtrsim_b(p,t+1)$; (iii) status quo: $D \sim_s (s,t)$ and $D \sim_b (b,t)$ for all t.

Notice that all preferences which are represented by utility functions of the form $U(p,t) = (p-s)\prod_{i=1}^{t}\delta_i$ fall into this category, but the lexicographic preferences which give time the secondary significance are ruled out.

The implementation game will be designed to take place over time and the design will include specification of the timing of the different decision nodes. We shall not identify one period with one move in the game, but assume that the design may prescribe a few moves to a single period.

Definition 3: A function P will be called <u>implementable (over time)</u> if there exists a game form with perfect information such that:

(1) Each node is dated: the date of the origin is 0 and, if node y comes after node x, the date attached to y is at least as late as the date attached to x.

(2) The terminal nodes are either outcomes of the type (p,t), where p in $[s_{min}, b_{max}]$ and t is the date of that terminal node, or the no-sale outcome D.

(3) For all s and b, and for <u>any</u> specification of the parties' time preference having the above properties, the unique SPE outcome is (P(s,b),0).

Notice that this definition extends definition 1 (in section 3) to refer to the added time dimension. Condition (1) describes how the time enters into the model. Condition (2) means that the implementation of the contract ends either with an exchange in which case the price paid by the buyer is received by the seller, or it ends with no sale in which case the parties do not make or receive any payment. Condition (3) gives the precise sense in which the contract implements exchange at P(s,b)--for all b and s it is the unique SPE outcome of the game laid out in the contract.

It is important to note that condition (3) holds for any specification of the parties' time preferences. This means that we regard a function P as implementable only if the design of the contract does not require knowledge of the time preferences, beyond the fact that they satisfy the three required properties[2].

We are interested in studying renegotiation-proof contracts in this context. This framework allows us to introduce a renegotiation proofness criterion which is not as extreme as the criterion of section 4.

Definition 4: A price function P is <u>renegotiation proof implementable (over time)</u>, if (1) it is implementable; (2) all terminal nodes are outcomes of the form (p,t); (3) if node y is an immediate successor of node x, then the date attached to it is equal to the date attached to x or to that date plus 1; (4)

for all s and b, any subgame and each period t, the SPE is Pareto efficient in the beginning of t (i.e., the SPE is efficient at any decision node at t which no other decision node at t precedes it on the path from the origin);

Condition (2) assures that there is no incentive to renegotiate the outcome after the implementation game is over. Notice that while in section 4 such a requirement assures that the outcomes are efficient when evaluated from any point in the game, this need not be the case here due to the presence of time costs. Therefore, conditions (3) and (4) are meant to supplement (2) by extending the renegotiation proofness requirement to other points in the game as well. Specifically, it assures that there is no incentive to renegotiate at certain points--the beginning of each period--throughout the game.

Note that the essence of the above conditions is that an implementation game is renegotiation proof, if after any history the SPE is almost Pareto Efficient in the sense that there is no possible exchange that would give each party a payoff that exceeds its expected SPE payoff by more than the value attributed by this party to one period.

Put differently, suppose that renegotiation is time consuming--the time it takes to tear up the old contract and negotiate a new one--and that it takes at least one time period to renegotiate a contract. Then, if a contract is renegotiation proof in the sense of this section, no attempt at renegotiation can be mutually beneficial since it is impossible that the post-renegotiation payoffs of both parties will be high enough to compensate for the time lost in renegotiation.

The following result shows that this criterion admits again the wide class of contracts which, as shown in section 3, are implementable in the absence of renegotiation proofness, but most of which were ruled out by the renegotiation proofness criterion of section 4.

Proposition 2:

Any function P, $s<P(s,b)<b$, which is increasing in s and b is renegotiation proof implementable over time.

Proof:

Consider the following game in extensive form.

Phase 1:

The seller announces a valuation v_S in S.
The buyer may challenge the seller or announce a valuation v_B in B.
If the buyer challenges, the game continues to Phase 2-0.
If the buyer announces v_B, the seller may agree or challenge.
If the seller challenges, the game will continue to Phase 3-0.
If the seller agrees to the buyer's announcement, $P(v_S,v_B)$ will be implemented.

Phase 2-t: A Bargaining Game

The seller makes a price offer p.
The buyer either accepts the offer and p is implemented or rejects it and makes a counter-offer $q<v_S$.
The seller either accepts the offer and q is implemented or rejects it and the game continues to Phase 2-(t+1).

Phase 3-t: A Bargaining Game

The buyer makes a price offer p.
The seller either accepts the offer and p is implemented or rejects it and makes a counter-offer $q>v_B$.
The buyer either accepts the offer and q is implemented or rejects it and the game continues to Phase 3-(t+1).

The decision nodes described in Phase 1 are in period 0. The nodes of phase 2-t and 3-t are in period t.

To every infinite path of the game we attach the outcome D.

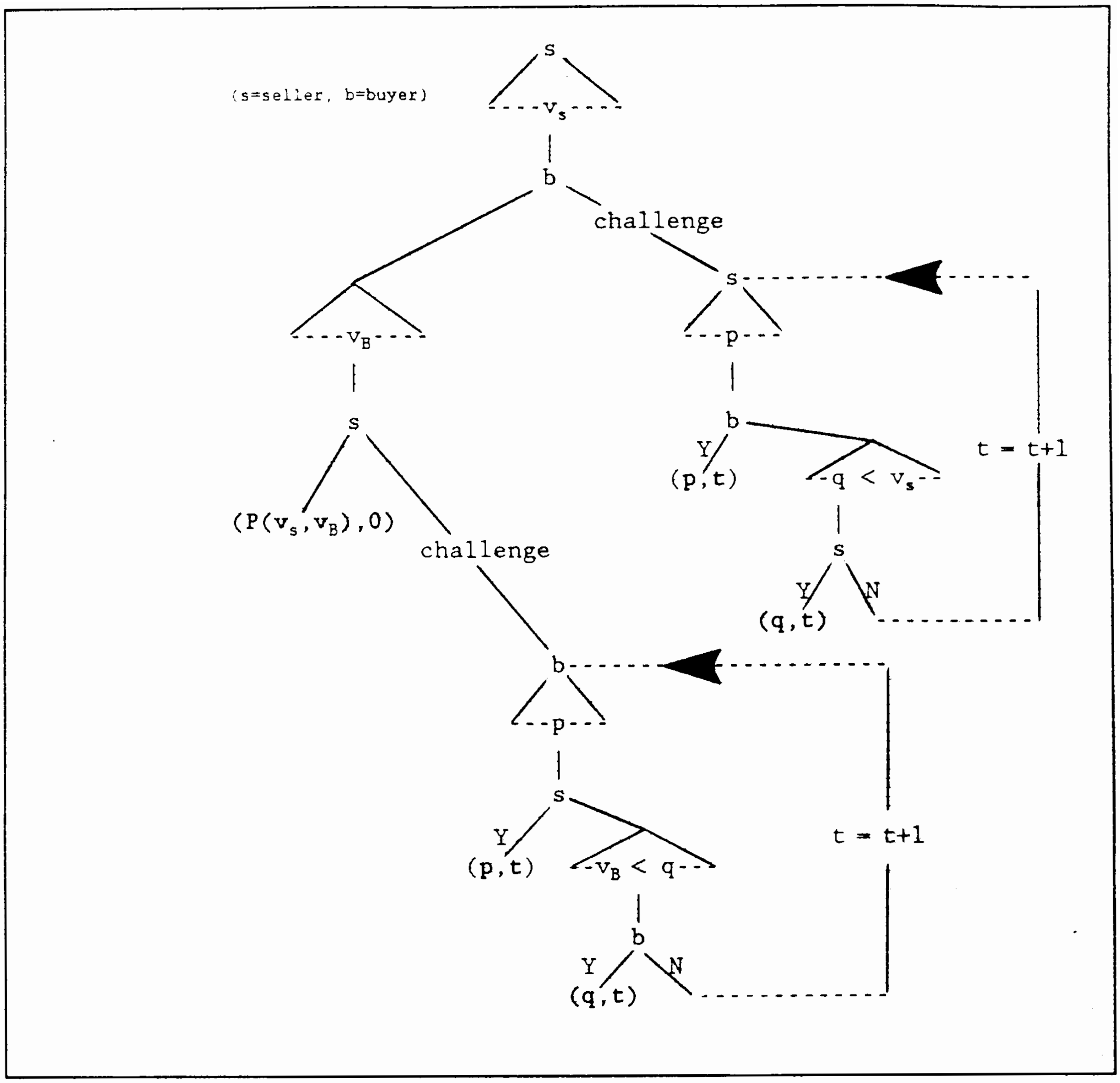

Figure 2

The proof that for b>s the game has a unique SPE is based on a similar idea to that of proposition 0.

Step 1: Consider a subgame that starts at the beginning of phase 2. If $v_S>s$, the game is effectively a bargaining game in which the buyer makes all the offers, since the buyer has the last word in each period and the restriction on the price offers is of no consequence in this case. With the time preferences allowed here (which display impatience, time-substitution and the status-quo properties), the unique SPE outcome in a bargaining game in which one side makes all the offers is such that this side appropriates all the surplus. Therefore, if $v_S>s$ the SPE outcome of this subgame is (s,0). Similarly, if $v_S \leq s$, the game is effectively a bargaining game in which the seller makes all the offers, since the restriction on the buyer's offers to be below v_S makes them irrelevant and therefore the SPE outcome is (b,0).

Consider next a subgame that starts at the beginning of Phase 3. By complete analogy, if $v_B<b$, the SPE outcome will be (b,0) and, if $v_B \geq b$, the SPE outcome will be (s,0).

Step 2: Suppose that the seller's announcement is $v_S=s$ and consider the continuation. If $v_B=b$, the seller will not challenge the buyer since the seller prefers (P(s,b),0) to (s,0), which by step 1 will result from challenging; if $v_B<b$ the seller will challenge the buyer and by step 1 the outcome will be (b,0); if $v_B>b$ the seller will not challenge and the outcome will be $(P(s,v_B),0)$ which by the monotonicity of P is worse for the buyer than (P(s,b),0).

Therefore, if the seller announces $v_S=s$, the buyer's equilibrium response is to announce $v_B=b$ and the SPE outcome is (P(s,b),0).

Step 3: Suppose that the seller announces $v_S>s$. As argued in step 1, by challenging the buyer will achieve the outcome (s,0). If the buyer does not challenge, the outcome will be either $(p(v_S,v_B),0)$ or the game will continue to phase 3 in which the seller will never agree to a price below s. Since $p(v_S,v_B) > v_S > s$, the buyer's best response is to challenge.

Step 4: If the seller announces $v_S<s$, then the buyer can announce $v_B=b$ and

the outcome will be $(P(v_S,b),0)$. By the monotonicity of P this price is lower than $P(s,b)$.

It follows from steps 2,3 and 4 above that the only SPE outcome of this game is $(P(s,b),0)$.

QED

The proposition shows that requiring a contract to be renegotiation-proof is not as restrictive as it might seem from the approach described in the previous section. When the criterion of renegotiation-proofness is weakened as would be sensible if the process of renegotiation itself is a time consuming activity, then a wide class of price functions are implementable.

One sense in which this proposition is strong is that <u>one</u> mechanism implements a price function P for <u>all</u> time preferences, as long as they reflect some degree of impatience. That is, given that time is costly, the design of a contract which is implementable and even in a renegotiation-proof manner does not require any further information about the time preferences.

Note that the criterion for renegotiation proofness is that the SPE is efficient at the beginning of each time period, but not necessarily at each node. If we required efficiency at each node, we would be back with the renegotiation proofness criterion of section 4, and nothing would be gained from the added structure. More precisely,

<u>Claim:</u> Modify part (4) of definition 4 to read:

"for all s and b, and any subgame (<u>not</u> only those which start at the beginning of a period), the SPE is Pareto efficient". Then the only functions P which are renegotiation-proof implementable over time are the constant ones.

Proof: Consider a function P which is renegotiation-proof implementable over time, according to the modified definition. Consider the case in which the time preferences of all seller and buyer types are represented by the utility functions $(p-s)\delta^t$ and $(b-p)\delta^t$ respectively, where $\delta<1$. It is sufficient to show that the SPE that induces the outcome $(P(s_{max},b_{min}),0)$ is a SPE for all (s,b). Assume that there is a subgame for which one of the agents, say seller s, can deviate profitably. By assumption, the SPE outcome in this subgame is Pareto efficient (for s_{max} and b_{min}), which means that this outcome is an exchange at some price p in this period. The profitability of the deviation by seller s means that he can achieve an exchange for p' after some t periods' delay such that $(p'-s)\delta^t>(p-s)$. However, if this inequality holds for s, it also holds for s_{max} so that this pair of strategies may not be a SPE for s_{max} and b_{min}.

QED

6. Discussion

Costly Renegotiation

One may argue that the implementation game of section 5 is not fully renegotiation-proof, even though it passes the criterion of definition 4. This is because there exist points in the game, not at the beginning of periods, at which the SPE is not efficient. At such points the parties would have an incentive to renegotiate to an efficient outcome. However, if renegotiation is costly, this criterion is immune to such criticism. Specifically, suppose that the cost of the renegotiation process is in it being time consuming. If the single time period of the implementation game is shorter than the amount of time needed to renegotiate the contract, then a contract that passes the criterion of section 5 is renegotiation-proof in the following sense. At any node of the implementation game, including those in which the SPE is inefficient, it will not be profitable for both parties to renegotiate. This is because any node is at most one time period away from an efficient node and this implies that any alternative agreement which will be implemented after a delay of at least one period cannot be preferred by both parties to the continuation of the game.

Of course, if renegotiation involves other costs instead of the time costs, a similar point can be made. As long as the costs of renegotiation are larger than the costs of delaying the implementation game for one period, renegotiation is prevented since it may not be profitable for both parties.

Why Renegotiation-Proofness?

We have yet to explain why the notion of a renegotiation-proof contract is useful for investigating the theoretical limits of contracting with non-verifiable information. That is, to explain the sense in which price functions which are implementable by renegotiation-proof contracts are the only ones that can at all be implemented in an environment that does not prohibit renegotiation. To think about this question, let us return to the world of section 4 (before the introduction of the time dimension), where renegotiation-proofness is identified with efficiency of the SPE in all subgames. Recall that a contract which is not renegotiation proof in that world is such that there are certain nodes of the implementation game and certain types s and b for which the outcome prescribed by the contract will be inefficient. This implies that the actual outcome will be achieved via renegotiation after the implementation game ends. Suppose that the function g(s,b,y) describes the results of the renegotiation between s and b after they reached the terminal node y. Given such a bargaining function, one may study the full consequences of a contract, even if it is not renegotiation-proof. The parties to such a contract will simply act as if the outcome at node y is the (possibly) type dependent outcome g(s,b,y) rather than the outcome prescribed by the contract. This approach is taken by Maskin and Moore(1986), Green and Laffont(1988) and Aghion, Dewatripont and Rey(1989), who investigate the scope of implementation in the presence of exogenously given bargaining function. Notice that since the bargaining function g is allowed to depend on s and b, there are price functions which are implementable given some function g, but are not implementable in a renegotiation-proof manner in the sense of the present paper.

Thus, it may appear that there is no reason to restrict attention to price functions P(s,b) which are implementable in a renegotiation-proof manner. This is because other functions which are implementable by non-proof contracts, given the appropriate renegotiation technology, may seem relevant as well. However, if we follow the logic of this reasoning further, this statement will seem questionable. If one believes that the bargaining process summarized by the function g(s,b,y) can be described as a game in extensive form, then it should be possible to complete a non-renegotiation proof contract by replacing node y with the bargaining game that results in g(s,b.y). (If different types engage in different bargaining games, the contract has to include the game that decides what bargaining games s and b will follow and so on.) Now, by the same reasoning the bargaining game itself should also be renegotiation-proof in the sense that inefficient terminal nodes will be renegotiated to efficient ones. Therefore, the compounded contract has to be renegotiation-proof in the sense of the present paper.

Thus, focusing on renegotiation-proof contracts in the sense of this paper is not an arbitrary restriction on the scope of the discussion. One may not suppose, for example, that the limitations on what can be implemented via renegotiation-proof implementation of the type considered in section 4 will be alleviated if we considered open ended contracts that leave for some types room to renegotiate. Once we recognize that any bargaining games that take place outside the contract can, in principle, be included in the contract, then proposition 1 characterizes all the price functions that can at all be implemented in an environment in which inefficient outcomes are renegotiated.

How Renegotiation Proofness Affects the Consequences of Non-Verifiability

The fact that the information of the parties to a contract is not verifiable to a third party is a form of imperfection. As other imperfections in the use of information, it could have real effects on the allocation of resources. For example, if one of the parties has to make some investment before the valuations are realized, it is possible to construct examples in which the contract has to condition the outcomes on the information in order to create incentives that will assure the proper magnitude of the investment. But when the information is non-verifiable, a contract that conditions on it may not be enforceable, and in the absence of such a contract the resulting investment might be inefficient.

Obviously, when the possibility of renegotiation exists so that the relevant contracts are the renegotiation-proof ones, the inefficiency/under-investment problem pointed out above may become more pronounced (see Green and Laffont for a discussion of this issue). For concreteness, consider the following example. $S=\{0\}$, $B=\{1,2\}$ and suppose that ex-ante there is equal probability for anyone of the buyer's types. Assume that in order to produce the unit the seller has to invest a sum of 1.2 before the buyer's valuation is determined. Recall that the renegotiation- proof contracts in the sense of proposition 1 admit only constant price contracts, which do not use any of the information. Therefore, if that is the right renegotiation-proofness restriction, the seller's revenue cannot exceed 1 and, hence, in the first place the seller will not invest in the production and inefficiency will result. The implication of Section 5 and proposition 2 for this example is that the inefficiency derived here owes to the too powerful renegotiation-proofness criterion. If renegotiation is time consuming or otherwise costly to the extent that the criterion of section 5 is appropriate, then the set of relevant contracts is richer and hence the potential inefficiency problem seems less severe. Here, the parties could have a renegotiation-proof contract which implements $P(0,1)=0.9$ and $P(0,2)=1.7$. This contract alleviates the inefficiency by making the investment of 1.2 profitable.

The Role of Time

This paper recognizes and exposes the important role that time may have in the design of mechanisms for the enforcement of contracts. Two properties of time are used here: that time is costly and that its passage is irreversible. Of course, the analysis does not necessitate the incorporation of a real time dimension into the contract. The effects of time can be mimicked by imposing other costs on the parties. Nevertheless, the real time dimension is of primary importance in this problem, since other forms of "burning" resources may not be commonly observed in practice. Furthermore, it is natural to think of the renegotiation process itself as involving time and therefore to design a contract that takes into account the time dimension of the renegotiation processes.

An Alternative Interpretation

Throughout this paper we interpreted the implementation problem under consideration as a contract design problem. An alternative interpretation of this analysis is that it studies the properties of a family of games. For example, the analysis of section 4 and proposition 1 can be viewed as a statement on the family of games with perfect information, parameterized by the seller's and buyer's reservation values, all of whose terminal nodes are exchanges at prices from the interval $[s_{min}, b_{max}]$ and for all s and b such that $s<b$, there is a unique SPE payoff. What proposition 1 tells us about this family of games is that the outcome function (the outcome as a function of the parameters s and b) is constant over all (s,b) combinations such that $b>s$.

From this point of view proposition 1 seems more interesting than 2, since it gives a tighter characterization of the outcome function of a family of games, while from the point of view of contract design proposition 1 seems as a negative result and proposition 2 is the more informative one.

End-Remarks

The basic seller-buyer scenario which we analyze has a rather special structure. Two major characteristics of this scenario are that (a) the parties preference rankings of all outcomes but the no-sale outcome are diametrically opposed; and (b) all types of seller (or buyer) have almost identical preferences and differ only with respect to how they rank the no-sale outcome vis-a-vis others. In addition, we focus on the class of contracts that are ex post individually rational in the sense that $s<P(s,b)<b$.

These limitations do not allow to speculate much beyond the boundaries of the above discussion of the theoretical possibilities of contracting with non-verifiable information. Nevertheless, we believe that the idea of explicitly including the time dimension in an implementation problem, in general, and in such a problem with renegotiation, in particular, has validity beyond confines of the present model.

Appendix I

Relaxing the assumption that S and B are disjoint

The above analysis was conducted for the case in which for any realization of s and b we have $s<b$ so that trade is always efficient. In what follows we comment on how the above analysis and its results can be extended to accommodate the case in which S and B overlap.

To avoid discussion of degenerate cases, it is assumed that $b_{max} \geq s_{max}$ and $b_{min} \geq s_{min}$. We shall be interested in contracts which for each pair s,b such that $s<b$ implement a price function P , $s<P(s,b)<b$, and for pairs s,b such that $s>b$ result in no trade.

Consider first the material of section 3. Definition 1 will be modified as follows to accommodate the possibility of $b<s$.

Definition 1': A price function P(s,b) will be called implementable if there exists a game form with perfect information such that: (1) all its terminal nodes are either (exchanges for) prices in the interval $[s_{min}, b_{max}]$, or the no-sale outcome D; (2) for all s,b such that $b>s$, the unique subgame perfect equilibrium outcome is an exchange at the price P(s,b); (3) for s,b such that $s \leq b$, the subgame perfect equilibrium payoffs are non positive.

Note that definition 1' differs from definition 1 in the added condition (3). We think of the game form as a procedure fixed by a contract. This procedure can be activated after the values s and b are realized, if at least one of the parties wants. If both choose not to activate the contract they just not trade. Condition (3) means that when there is no surplus to split, no party has an incentive to activate the procedure laid out in the contract.

Proposition 0 does not change. The game described in the proof already satisfies condition (3) so that the only addition to the proof is that in the announcement stage if $v_B \leq v_S$ the outcome is D.

Consider next the material of section 4. Let us leave definition 2 as it is except that condition (3), that the SPE of all subgames is efficient, will be required only for realizations such that s<b.

<u>Proposition 1'</u>:

If $b_{min} < s_{max}$ then there is no renegotiation proof implementable price function P.

<u>Proof</u>: Suppose that the function P(s,b) is such that, for all b>s, b>P(s,b)>s is renegotiation-proof implementable. It follows from proposition 1 that $P(s,b_{max})$ is constant for all s and that $P(s_{min},b_{max}) \geq s_{max}$. Also, $P(s_{min},b)$ is constant for all b and $P(s_{min},b_{max}) \leq b_{min}$.

Thus, if $b_{min} < s_{max}$ the above leads to contradiction and there exists no such renegotiation-proof implementable function.

QED

Thus, although the Pareto efficiency of the SPE in all subgames is imposed only for realizations such that s<b, there is no renegotiation proof function.

Finally, in order to consider the results of section 5 we have to modify definitions 3 and 4 as follows.

Definition 3': A function P will be called <u>implementable (over time)</u> if there exists a game form with perfect information such that:

(1) Each node is dated: the date of the origin is 0 and, of course, if node y comes after node x, the date attached to y is greater than or equal to the date attached to x.

(2) The terminal nodes are either outcomes of the type (p,t), where p in $[s_{min}, b_{max}]$, or the no-sale outcome D.

(3) For all s and b such that b>s, the unique SPE payoff is P(s,b), for <u>any</u> specification of the parties' time preference having the above properties.

(4) For any s and b such that s≥b, all SPE payoffs are non-positive.

Notice that this definition simply differs from definition 3 in the added condition (4) which requires that when there is no surplus to split, no party has an incentive to activate the procedure laid out in the contract.

Definition 4': A price function P is <u>renegotiation proof implementable (over time)</u>, if (1) it is implementable; (2) for b>s, any terminal node dated t is Pareto efficient at t (i.e., all terminal nodes are outcomes of the form (p,t)); (3) for all s and b, any subgame and each period t, the SPE is Pareto efficient in the beginning of t (i.e., the SPE is efficient at any decision node at t which no other decision node at t precedes it on the path from the origin); (4) for any s and b, each of the parties has a strategy which guarantees to it payoff zero.

Definition 4' differs from 4 in condition (4) which is added to guarantee that the case of s>b is irrelevant, by making sure that in this case the parties will not want to activate the procedure. This is because, with s>b, there is no possible exchange that will benefit both, and by condition (4) each party can guarantee at least zero to itself, so that no party has an incentive to activate the procedure.

Proposition 2 and its proof remain the same. One has only to notice that the game described in the proof already satisfies condition (4) of definition 4'.

Appendix II

A proof that the requirement that P is decreasing in s and b is a necessary condition for the implementability of P by a finite game.

Assume that the function P is implementable by a finite game and that $P(s,b)<P(s,b')$ for some $b>b'$. The following lemma shows that the implementation game between s and b' must have a SPE whose outcome is D or a price below $p(s,b)$. Thus, the implementation game between s and b' has two SPE outcomes, in contradiction to the implementability of P. Therefore, P must be non-decreasing in b and, analogously, it has to be non-decreasing in s.

Lemma: Let (f_{sb},g_{sb}) be a SPE for a finite game with perfect information between s and b, let $b'<b$ and let $O(f,g)$ denote the outcome of strategies f and g.

If $O(f_{sb},g_{sb}) = p^*$, $p^* > b$, then $\exists$ SPE for (s,b') s.t. $O(f,g) = p^*$

If $O(f_{sb},g_{sb}) = p^*$, $b' \leq p^* \leq b$, then $\exists$ SPE for (s,b') s.t. $O(f,g) = D$ or $O(f,g) = p \leq p^*$

If $O(f_{sb},g_{sb}) = p^*$, $s \leq p^* < b'$, then $\exists$ SPE for (s,b') s.t. $O(f,g) = p$, $s \leq p \leq p^*$

If $O(f_{sb},g_{sb}) = p^* < s$, then $\exists$ SPE for (s,b') s.t. $O(f,g) = p^*$

If $O(f_{sb},g_{sb}) = D$, then $\exists$ SPE for (s,b') s.t. $O(f,g) = D$

Proof: By induction on the diameter of the game. It is obvious that the statement of the lemma is true for games of diameter zero. Assume that it holds for all finite games of diameter less than L, consider a game of diameter L and notice that all subgames of such a game are of diameter less than L.

Consider the different subgames $G_1, \ldots, G_k$ which follow the first move in the game. Let $(a_1, \ldots, a_k)$ denote their outcomes according to (f_{sb}, g_{sb}). By the inductive assumption each G_i has a SPE outcome a_i' in the game between s and b', where a_i' is related to a_i as described in the lemma. Define the strategies (f,g) as follows. The first move of the player who controls the root of the game is the one that selects the subgame G_i such that a_i' is the best outcome for this player from the set $\{a_1', \ldots, a_k'\}$ and the continuation is such that in each subgame G_i the strategies f and g coincide with the SPE strategies for s and b' which yield the outcome a_i', $i=1, \ldots, k$. Notice that (f,g) is a SPE in the entire game between s and b'.

Now suppose that $O(f_{sb}, g_{sb})$ is a_j and O(f,g) is a_i' and let us verify that they are related as required by the lemma.

Assume first that the buyer moves at the root of the game. If a_j is a price and $a_j > b$, then all outcomes in $\{a_1, \ldots, a_k\}$ must be prices above b and by the inductive hypothesis $a_m' = a_m$ for all m and therefore the minimal price in $\{a_1, \ldots, a_k\}$ is the same as the minimal price in $\{a_1', \ldots, a_k'\}$. If a_j is a price and $b' \leq a_j \leq b$, then by the inductive hypothesis either $a_i' \leq a_j' \leq a_j$ or $a_i' = D$. If $s \leq a_j < b'$, then for all n, $a_n = D$ or $a_n \geq a_j > s$. By the inductive hypothesis a_j' is also a price and $s \leq a_j' \leq a_j < b'$ and, for all n, $a_n' \geq s$ or $a_n' = D$. Hence a_i' is a price and $s \leq a_i' \leq a_j' \leq a_j \leq b'$. If $a_j < s$ then for all n, $a_n = D$ or $a_n \geq a_j$. By the inductive hypothesis, $a_j' = a_j$ and for all n $a_n' = D$ or $a_n' \geq a_n$. Hence, $a_i' = a_j$. Finally, if $a_j = D$, then all outcomes in $\{a_1, \ldots, a_k\}$ are either prices above b, or D, and hence by the inductive assumption all outcomes in $\{a_1', \ldots, a_k'\}$ are also either prices above b or D implying that $a_i' = D$. Thus, the constructed SPE has the required property.

The remaining case is that the seller has the first move. The argument in this is similar to the above case, and we leave it to the reader. QED

References

Abreu, D. and A. Sen (1987), "Virtual Implementation in Nash Equilibrium," mimeo.

Aghion, P., Dewatripont, M. and P. Rey (1989), "Renegotiation Design Under Symmetric Information," mimeo.

Glazer, J. and Ma, A.(1988) "Efficient Allocation of a Prize--King Solomon's Dilemma," mimeo Boston University.

Green, J. and Laffont, J. (1987), "Renegotiation and the Form of Efficient Contracts," mimeo MIT.

Hart, O and Moore, J. (1988),

Matsushima, H. (1988), "A New Approach to the Implementation Problem," JET, 45, 128-144.

Moore, J. and Repullo, R. (1988), "Subgame Perfect Implementation," Econometrica, 56, 1191-1120.

Maskin, E. and Moore, J. (1986), "Implementation of Renegotiation," mimeo.

Notes

1. We adopt throughout the convention that "increasing" means weakly increasing.

2. Compare with the virtual implementation ideas of Abreu and Sen (1987) and Matsushima (1988).