Renegotiation-Proof Implementation and Time Preferences

By Ariel Rubinstein and Asher Wolinsky*

This paper explores how the requirement that the implementation of contracts be renegotiation-proof affects the set of contracts that can be implemented in a seller – buyer scenario in which the information regarding the agents' valuations is nonverifiable. This paper explicitly adds a time dimension to an implementation problem and introduces a natural criterion of renegotiation-proofness for the case of time-consuming renegotiation. The main insight gained is that the addition of the time dimension enlarges significantly the set of contracts that can be implemented in a renegotiation-proof manner. (*JEL* C70, D23)

This paper explores how the requirement that the implementation of contracts be renegotiation-proof affects the set of contracts that can be implemented in a seller–buyer scenario in which the information regarding the agents' valuations is nonverifiable. The paper's main contributions are 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.

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 nonverifiable (i.e., not observable to third parties, such as a court).[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 $P(s,b)$ that can be so implemented by a contract and examine how this set is affected by the requirement that the contract be renegotia-

*Department of Economics, Tel-Aviv University, Tel-Aviv, Israel, and Department of Economics, Northwestern University, Evanston, IL 60208, respectively. Rubinstein 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 begun. Wolinsky gratefully acknowledges research support from NSF grant SES-9009509. We thank Dilip Abreu, Matt Jackson, Albert Ma, Steve Matthews, John Moore, David Pearce, Rob Porter, and three anonymous referees for useful comments. We owe special thanks to a referee who pointed out important flaws in an earlier draft.

[1]The distinction between verifiable and observable information was, we believe, first made by Sanford Grossman and Oliver Hart (1986).

tion-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 will 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. However, before we proceed to the description of the results, let us point out why the scenario and the questions we analyze to demonstrate these ideas are in themselves interesting objects for economic analysis. The following discussion follows Jean Tirole (1986).

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 each of the buyer types. Assume that in order to produce the unit the seller has to invest 1.2, 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 in 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 underinvestment amounts, therefore, to inquiring whether or not it is possible to implement other price functions, such as the one that prescribes prices 0.9 and 1.7 to the low- and high-valuation buyers, respectively. If the situation is such that the parties are free to renegotiate, the relevant question is whether other such price functions can be implemented by a renegotiation-proof contract, when the information is not verifiable. Of course, by identifying the set of all price functions implementable by a renegotiation-proof contract, we may address a class of such questions at once.

We restrict attention to the simplest case in which, for all realizations of s and b, $b > s$. In the first case considered, the possible outcomes that 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 that are nondecreasing in s and b and satisfy $b > P(s,b) > s$. 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 agree mutually to scrap the mechanism, they will probably not put up with inefficiency and instead will negotiate a new outcome. Therefore, if agents see through the contract and anticipate renegotiation, they will not necessarily be deterred by the no-sale outcome, and hence the contract might not achieve the desired outcomes.

This criticism motivates the work of Tai-Young Chung (1988), Jerry Green and Jean-Jacques Laffont (1988), Hart and John Moore (1988), Eric Maskin and Moore (1988) and Philippe Aghion et al. (1989). They respond to it by looking at contracts that take into account the negotiated outcome that will follow an inefficient one. We follow here a somewhat different approach and look at contracts that 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 subgame-perfect equilibrium (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 that is preferred by both agents. 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. The reason is that this limited subset may not contain sufficiently rich contracts which are required under certain circumstances to provide the right incentives for, say, investment that 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 a time dimension and that, after the implementation game is over, the parties turn to renegotiating inefficient outcomes. However, the above approach leaves this dimension unmodeled 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 explicitly adding 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 timetable of the different steps in the execution of the contract. The two features of the time dimension that 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 invoked here requires that, after any possible history, the subgame-perfect equilibrium outcome is not Pareto-dominated by any possible outcome one period hence. That is, if one period is the minimum amount of time required to renegotiate a contract, then at no point will both parties find it mutually advantageous to renegotiate.

It turns out that with this notion of renegotiation-proofness, all contracts that specify trade at the price $P(s,b)$, where $P(s,b)$ is nondecreasing 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. This is in sharp contrast to the case in which only efficient outcomes are considered to be renegotiation-proof.

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 Rafael Repullo (1988).

I. The Model

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 finite sets S and B.[2] Let s_{max}, b_{max}, s_{min}, and b_{min} denote the maxima and minima of S and B, respectively. Throughout this paper we assume that $b_{min} > s_{max}$. This means that, for all possible realizations of (s,b), there are gains from trade.[3] When

[2]This assumption is not essential in the following sense. It is used only in the proof of Proposition 2, and as we note there, we can prove a very similar result for the case in which S and B are intervals.

[3]If the sets S and B are allowed to overlap (i.e., $s_{max} > b_{min}$), the main consequence for this analysis is that the set of *ex post* efficient outcomes would depend on the state (s,b): trade would be *ex post* efficient if and only if $s < b$. This fact would introduce further difficulties into the subsequent analysis of renegotiation-proofness. In appendix I of our working paper (Rubinstein and Wolinsky, 1990), we present one way

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 in which, 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 that are strictly *ex post* individually rational [i.e., satisfy $s < P(s,b) < b$]. This means that we restrict attention to situations in which the parties cannot enforce outcomes that are not *ex post* individually rational (for some reasons which are left unmodeled here).

Definition 1: A price function $P(s,b)$ will be called *implementable* if there exists an extensive game form with perfect information such that: (i) all its terminal nodes are either (exchanges for) prices, or the no-sale outcome D; (ii) 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 that 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.

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.

II. 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 Jacob Glazer and Albert Ma (1989).

PROPOSITION 0: *Any function* P, $s < P(s,b) < b$, *that is nondecreasing in* s *and* b *is implementable.*

PROOF:

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

Stage 1: The Announcement Stage.—The seller announces a number v_S in S ("declares his valuation"). The buyer challenges the seller or announces a pair (v_B, v'_S), where v_B is in B, v'_S is in S, and $v'_S \geq v_S$ ("declares his own valuation and his upward correction for the seller's declared valuation"). If the buyer challenges the seller, the game continues to stage 2. If the buyer chooses (v_B, v'_S), 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

of extending the definitions and the results to the case in which S and B may overlap.

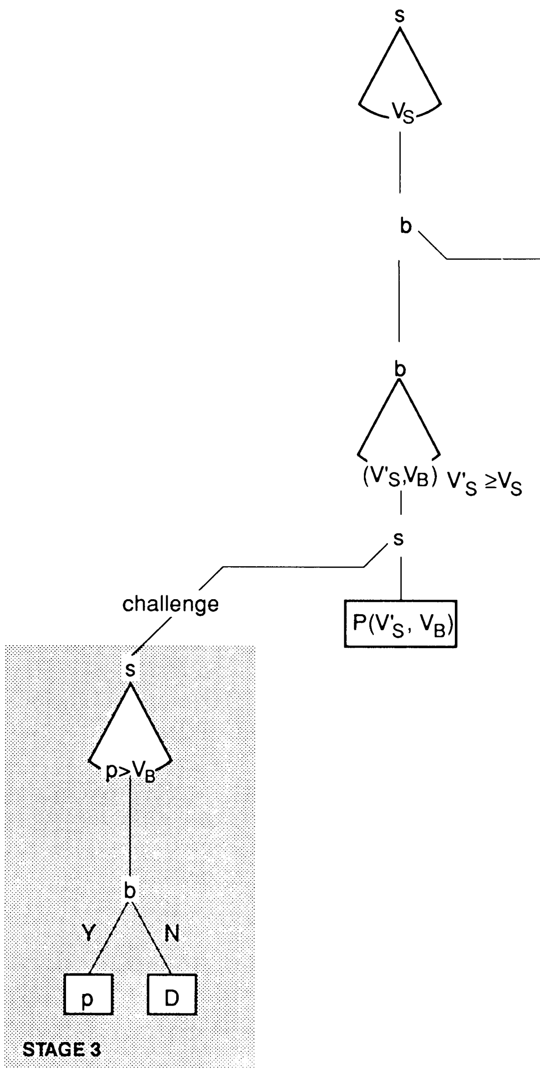

FIGURE 1. THE IMPLEMENTATION GAME OF PROPOSITION 0

Note: s = node controlled by seller; b = node controlled by buyer.

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).

This game is depicted in Figure 1. Let us verify that it implements P. First consider a subgame in which the buyer makes a "take it or leave it" offer (stage 2). If $v_S > s$ the SPE outcome is exchanged for the price s, and if $v_S \leq s$, the SPE outcome is D. Similarly, for a subgame in which the seller makes a "take it or leave it" offer (stage 3), if $v_B < b$ the SPE outcome is an exchange for the price b, and if $v_B \geq b$ the SPE outcome is D.

Consider now the subgame that starts after the seller announced v_S. There are the following three cases.

Case 1: $v_S > s$.—If the buyer challenges, then the SPE outcome is an exchange at the price s; if he does not challenge and announces $v_B < b$ and any $v'_S \geq v_S$, the SPE outcome is an exchange at b; and if he announces $v_B \geq b$ and any $v'_S \geq v_S$, the outcome is $P(v'_S, v_B) \geq P(v_S, b) \geq P(s,b) > s$. Thus, in any SPE the buyer challenges the seller's announcement, and the outcome is an exchange at s.

Case 2: $v_S < s$.—In all SPE of such a subgame, the price cannot exceed $P(s,b)$, since the buyer can announce $v_B = b$ and $v'_S = s$.

Case 3: $v_S = s$.—If the buyer challenges, the SPE will be D. If the buyer announces $v_B = b$ and any $v'_S \geq v_S$, the seller will not challenge him in the following SPE, and the outcome will be an exchange at the price $P(v'_S, b) \geq P(s,b)$. In particular, if $v_B = b$ and $v'_S = s$, the outcome will be $P(s,b)$. If the buyer announces $v_B < b$ and any $v'_S \geq v_S$, 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 $v_B > b$ and any $v'_S \geq v_S$, then any SPE outcome in the continuation will be $P(v'_S, v_B) \geq P(s,b)$. Thus, after the seller announces $v_S = s$, any SPE outcome will be exchange for $P(s,b)$.

Since $P(s,b) > s$, which is the price the seller will get if $v_S > s$, and $P(s,b) \geq P(v_S, b)$ is the most the seller may get if $v_S < s$, it follows from the above three cases that the only SPE outcome is an exchange for $P(s,b)$, which the seller can achieve by announcing s.

The proof has already exposed the manner in which the implementation game works. Let us just call attention to the subtle role of the buyer's announcement of v'_S, which essentially gives him an opportunity to revise upward the seller's previous announcement of his valuation v_S. If the buyer did not have this option, in certain cases the seller would have an incentive to underreport his true cost.[4] To see this, suppose that $S = \{1,2\}$, $B = \{4,5\}$, $P(1,4) = 1.5$, $P(2,4) = 2.5$, $P(1,5) = 3$, and the true valuations are $s = 2$ and $b = 4$. Suppose that the implementation game is as above except that the buyer announces only v_B (rather than v_B and v'_S as above). Now, if the seller underreports by announcing $v_S = 1$, the buyer's best response is to overreport by announcing $v_B = 5$. This is because, by truthfully announcing $v_B = 4$, the seller will be made to choose between $P(1,4)$ and the outcome D which will result from challenging the buyer's announcement. Since $P(1,4) = 1.5 < 2 = s$, the seller's best response is to enforce D, and since $P(1,5) = 3 < 4$ the buyer's best response to $v_S = 1$ is $v_B = 5$. Thus, the game without the v'_S announcement, does not implement $P(2,4)$. The way in which this announcement avoids the above problem is in letting the buyer "revise" v_S upward by announcing $v'_S = 2$, leaving the seller with the choice between $P(2,4) = 2.5 > s$ and D, rather than the choice between $P(1,4) = 1.5 < s$ and D. Note also that the proof has not used the finiteness of S and B, and hence it holds without any modification for the case in which S and B are intervals.

Looked upon from the point of view of contract design, the result of Proposition 0 is optimistic, since it tells us that a wide range of price functions can be implemented. In particular, for the example from Tirole (1986) reported in the Introduction, one can implement the price function $P(1) = 0.9$ and $P(2) = 1.9$, which would facilitate efficient investment.

Although we shall not be concerned here with characterizing the exact set of implementable functions, let us point out that the sufficient condition on P, that it be nondecreasing in both arguments, is "almost" a necessary condition as well. In appendix II of the working paper (Rubinstein and Wolinsky, 1990), we show that any function P which is implementable by a finite game form (i.e., the number of nodes in the game tree is finite) is nondecreasing in s and b.

[4]We are grateful to an anonymous reader who pointed out this possibility.

III. Renegotiation-Proofness Is Identified with Efficiency

The proof of Proposition 0 relies on the possibility of enforcing the no-sale outcome, D: if the seller lies about the buyer's valuation, then the equilibrium outcome in the resulting subgame in stage 3 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 we should look at contracts that 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 it is implementable and the game form that implements it is such that, for all s and b, all nodes (including terminal ones), and every SPE, the SPE outcome in the subgame starting at that node is efficient.

In particular, this definition implies that all the terminal nodes are efficient (i.e., the no-sale outcome D cannot be reached after a finite number of moves). For finite game forms, this means that the outcome must be a trade; but for infinite game forms, which we do not exclude, this definition still allows one to identify an infinite path (of an infinite implementation game) with the outcome D.

This definition fits an environment in which renegotiation is costless and uninhibited in any way: any inefficiency at any node of the implementation game will presumably be instantaneously renegotiated away. As we have already mentioned, Chung (1988), Green and Laffont (1988), Hart and Moore (1988), Maskin and Moore (1988), and Aghion et al. (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 (see the discussion in Section V).[5]

The following proposition shows how this requirement of renegotiation-proofness reduces considerably the set of implementable price functions.

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. Let the strategy pair (f,g) be an SPE for the pair $(s_{\max}, b_{\min})$. Let (s,b) be any other pair of reservation values. We claim that (f,g) is an SPE for (s,b) as well. Consider any subgame. By the choice of (f,g) and the second condition of Definition 2, the outcome of the subgame when played according to (f,g) is efficient (i.e., it is a price p^*). Now, if seller s can deviate profitably, it means that either (i) he has a strategy which, when played against g, induces a price above p^* or (ii) he prefers D to p^* and has a strategy which, against g, induces the outcome D. In case (i), this deviation is obviously profitable for $s_{\max}$ as well, a contradiction. Case (ii) holds if $p^* < s < s_{\max}$, but then this deviation is also profitable for $s_{\max}$, a contradiction. Therefore, s cannot deviate profitably, and by an analo-

[5]Actually Chung (1988) and Aghion et al. (1989) take an intermediate approach. In their models, the renegotiation is not entirely exogenous: the contract may assign all the bargaining power in the renegotiation stage to one of the parties.

gous argument, b cannot deviate profitably either.

Thus, (f,g) is an SPE for types (s,b) as well. Since the implementability of P means that $P(s,b)$ is the *unique* SPE outcome in the game between s and b, the fact that (f,g) is an SPE for any types (s,b) implies that, for all (s,b), $P(s,b)=P(s_{\max},b_{\min})$.

Recall that Definition 2 does not completely exclude outcome D since it still allows infinite paths of infinite game forms. Note, however, that this only makes the proposition stronger: if this form of disagreement were also excluded, the result of Proposition 1 will hold a fortiori.

Observe 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.

IV. 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 $0,1,2,\ldots$, 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 over these outcomes extend the preferences over the basic outcomes. They are given by the utility functions $(p-s)\delta_s^t$ and $(b-p)\delta_b^t$ for seller type s and buyer type b, respectively; utility of zero is assigned by a seller and a buyer of any type to the no-sale outcome D.[6]

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 shall assume that the design may prescribe a few moves to a single period. Let $t(x)$ denote the date attached to node x.

A *dated game form* is an extensive game form with perfect information such that each node is dated: (i) the root's date is 0; (ii) if node y is an immediate successor of node x, then $t(y)=t(x)$ or $t(y)=t(x)+1$; and (iii) only a finite number of nodes on any single path from the origin have the same date. Condition (ii) means that there is a node in each period, even if no real decision is made at such a node. Its role will become clearer after we introduce the notion of renegotiation-proofness in Definition 4, below.

Definition 3: A function P will be called *implementable* (over time) if there exists a dated game form such that: (i) the terminal nodes are either outcomes of the type (p,t), where $p\geq 0$ and t is the date of that terminal node, or the no-sale outcome D; (ii) for all s and b, the unique SPE outcome is $(P(s,b),0)$.

Definition 3 extends Definition 1 (in Section I) to refer to the added time dimension. Condition (i) 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 (ii) gives the precise sense in which the contract implements $P(s,b)$: for all b and s, an immediate exchange for that price is the unique SPE outcome of the game laid out in the contract.

[6]We restrict attention to these time preferences for simplicity of the exposition. All arguments go through for a wider specification of time preferences (see Rubinstein and Wolinsky, 1990 p. 15).

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 III. Let $O(s,b,x,e)$ denote the SPE outcome in the subgame starting at node x when it is played by s and b according to the SPE e. If x is a terminal node, then $O(s,b,x,e)$ is the outcome corresponding to it.

Definition 4: A price function, P, is renegotiation-proof implementable (over time), if it is implementable by a dated game form that satisfies the following condition. For all s and b, any node x and each SPE e, there is no p such that both s and b prefer $(p,t(x)+1)$ to $O(s,b,x,e)$.

The essence of this definition 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, one period hence, that would give each party a payoff that exceeds its expected SPE payoff. Recall that the definition of a dated game form requires that there is at least one node in each period. This assures that the renegotiation criterion is applied to each period (i.e., the implementation game cannot impose long delays during which the parties are simply barred from renegotiating).

Put differently, suppose that renegotiation is time-consuming (the time it takes to scrap 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 successful, since the postrenegotiation payoffs of both parties cannot 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 that are implementable in the absence of renegotiation-proofness, as shown in Section II, but most of which were ruled out by the renegotiation-proofness criterion of Section III.

PROPOSITION 2: *Any function P, $s < P(s,b) < b$, that is nondecreasing in s and b is renegotiation-proof implementable over time.*

PROOF:

Define $\varepsilon = \min_{s,b}\{P(s,b)-s,\ b-P(s,b)\}$. Since S and B are finite, $\varepsilon > 0$. Let T be such that $\max_s\{(b_{\max} - s)\delta_s^T\} < \varepsilon$ and $\max_b\{(b - s_{\min})\delta_b^T\} < \varepsilon$, for all s and b. That is, ε is the minimum gains from trade for either buyer and seller; T is such that the discounted gains from trade T periods in the future are below ε, for all sellers and buyers. Consider the following extensive game form.

Stage 1: The Announcement Stage.—The seller announces a valuation v_S in S. The buyer challenges the seller or announces a pair (v_B, v'_S) where v_B is in B, and $v'_S \geq v_S$ is in S (i.e., declares his own valuation and his upward correction for the seller's declared valuation). If the buyer challenges, the game continues to stage 2-0. If the buyer chooses (v_B, v'_S), the seller may agree or challenge. If the seller challenges, the game will continue to stage 3-0. If the seller agrees to the buyer's announcement, $P(v'_S, v_B)$ will be implemented.

Stage 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 counteroffer $q(t)$: for $t \leq T$, $q(t) < v_S$; for $t > T$, $q(t)$ is unrestricted. The seller either accepts the offer [and $q(t)$ is implemented] or rejects it [and the game continues to stage 2-$(t+1)$].

Stage 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 counteroffer $r(t)$: for $t \leq T$, $r(t) > v_B$; for $t > T$, $r(t)$ is unrestricted. The buyer either accepts the offer [and $r(t)$ is implemented] or rejects it [and the game continues to stage 3-$(t+1)$]. The decision nodes described in stage 1 are in period 0. The nodes of stage 2-t and 3-t are in period t.

To every infinite path of the game we attach the outcome D. The proof that for $b > s$ the game has a unique SPE outcome

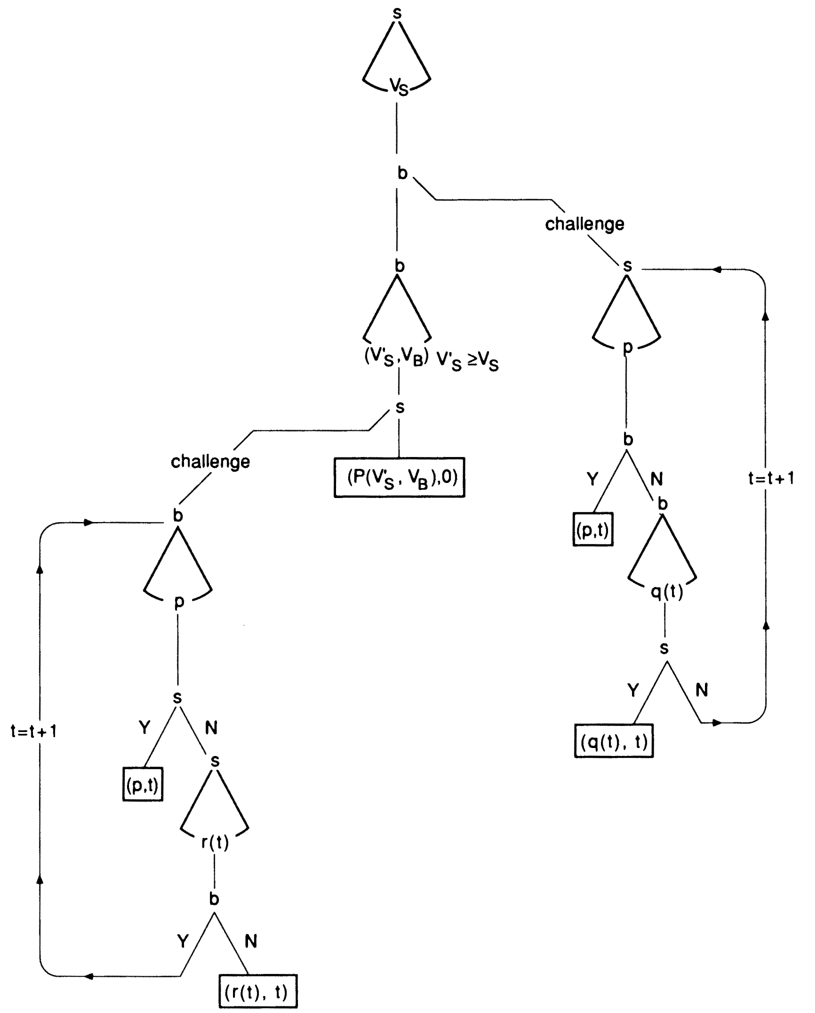

FIGURE 2. THE IMPLEMENTATION GAME OF PROPOSITION 2

Notes: s = node controlled by seller; b = node controlled by buyer. For $t \leq T$, $q(t) < V_S$; for $t > T$, $q(t)$ is unrestricted. For $t \leq T$, $r(t) > V_B$; for $t > T$, $r(t)$ is unrestricted.

is based on a similar idea to that of Proposition 0. Notice that the game's tree, as depicted in Figure 2, differs from the game of Figure 1 only in the subgames starting in stages 2 and 3. Therefore, let us first show that the subgames that start at the beginning of stages 2 and 3 have unique SPE outcomes and then establish that the proof of Proposition 0 continues to hold after substituting these outcomes for those of stages 2 and 3 in that proof.

Consider a subgame that starts at the beginning of stage 2. If $v_S > s$, then the subgame that starts at period $T+1$ is effectively a bargaining game in which the buyer makes all the offers, since the buyer has the last word in each period and there is no restriction on the buyer's price offers. With the time preferences allowed here, the unique SPE outcome in this subgame is immediate agreement on the price s [i.e., $(s, T+1)$]. Proceeding by backwards induction, the unique SPE outcome in the subgame starting at any $t \leq T$ is (s,t). Therefore, if $v_S > s$ the unique SPE outcome of the subgame that starts at the beginning of stage 2 is $(s,0)$.

If $v_S \leq s$, the game is effectively a bargaining game in which the seller makes all the offers up to period T, and from $T+1$ on the buyer makes all the offers. This is because the restriction on the buyer's offers to be below v_S up to time T makes them irrelevant, and the fact that from $T+1$ on the buyer has the last (unrestricted) word makes the seller's offers from that time on irrelevant. Therefore, the unique SPE outcome at $T+1$ is $(s, T+1)$. Proceeding by backwards induction, if $v_S \leq s$, the unique SPE outcome of the subgame that starts at the beginning of stage 2 is $(q,0)$ such that $b - q = (b-s)\delta_b^{T+1}$. By the choice of T, we know that this q satisfies $b - \varepsilon < q < b$.

Consider next a subgame that starts at the beginning of stage 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 $(r,0)$, where $s < r < s+\varepsilon$.

Now, what distinguishes these SPE outcomes from those of stages 2 and 3 in Proposition 0 are the cases in which the buyer challenges after $v_S \leq s$ or the seller challenges after $v_B \geq b$. In the former case, the SPE outcome of stage 2 of Proposition 0 was D, while here it is $(q,0)$ where $b - \varepsilon < q < b$. Analogously, in the latter case, the SPE of stage 3 of Proposition 0 was D, and here it is $(r,0)$ such that $s < r < s+\varepsilon$. Thus, in order to invoke the proof of Proposition 0, it remains only to verify that here too the buyer does not challenge after $v_S \leq s$ and the seller does not challenge after $v_B \geq b$ (in the case of $v'_S = s$), as is the case in the SPE of Proposition 0. After $v_B \geq b$, and provided that $v'_S = s$, the seller will get $r < s + \varepsilon$ by challenging and $P(v'_S, v_B) \geq P(s,b)$ by not challenging. The choice of ε assures that $P(s,b) - s > \varepsilon$; hence, $P(v'_S, v_B) - s > \varepsilon$, and the seller will not challenge. An analogous argument establishes that after $v_S \leq s$ the buyer will not challenge either.

Finally, observe that this game form satisfies the conditions of Definition 4. We have already shown that the SPE is efficient in the subgames of stage 1 and in the subgames that start at stages 2-0 and 3-0. It can be verified in the same manner that the SPE outcomes in the subgames that start at the beginnings of stages 2-t and 3-t are immediate agreements at t and hence are efficient at those nodes.

Before proceeding let us comment briefly on some features of the game form constructed above and how different assumptions were used in it. First, the role of the limit T on the duration of the restriction on offers is to assure uniqueness of the SPE in the subgames that follow the buyer's challenge of $v_S > s$ or the seller's challenge of $v_B < b$. If such a time limit were not placed, so that in stage 2, for example, the buyer's offers were always restricted to be below v_S, then for sufficiently large δ there will be other SPE as well. For example, when δ is large enough, there is a price $p > v_S$ that can be supported as an additional SPE outcome by the following strategies: the seller always offers $p > v_S$, and as long as he has not deviated from p, rejects all the buyer's offers; the buyer always offers s and agrees only to p; if the seller ever deviates from p, they switch to the equilibrium that supports s.

Second, in order to come up with a finite T that works for all types, there has to be some $\varepsilon > 0$ such that, for all s and b, $P(s,b) - s > \varepsilon$ and $b - P(s,b) > \varepsilon$. This follows immediately from the finiteness of S and B, and this is where we used that assumption. If we wanted to assume instead that S and B are intervals, then Proposition 2 would state that any function P that is nondecreasing in s and b, and which for some $\varepsilon > 0$ and all s and b satisfies $s + \varepsilon < P(s,b) < b - \varepsilon$, is renegotiation-proof implementable over time.

Third, the fact that more than one move can be made within one period plays a role in the construction. When, for example,

$v_S > s$ this feature effectively turns stage 2 into a bargaining game in which the buyer makes all the offers and hence appropriates all the surplus. If, alternatively, each move required the passage of costly time, the first T periods of stage 2 would be a standard alternating-offers game, and the unique SPE would not award all the surplus to the buyer.

This 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.

The game form on which the proof is based has an additional desirable property: *one* mechanism implements a price function P for a wide range of specifications of the time preferences.[7] The proposition can therefore be strengthened as follows.

COROLLARY: *For any $\bar{\delta} < 1$ and any function P as described in Proposition* 2, *there is one game form which implements P in renegotiation-proof manner for all specifications of the time preferences such that δ_s and δ_b are bounded from above by $\bar{\delta}$.*

The proof of the corollary follows immediately from noting that, in the proof of Proposition 2, the magnitude of the discount factors matters only through the determination of T, which can be determined in reference to $\bar{\delta}$ alone. The interest of this observation is that, given that time is costly, the design of a contract that is implementable and even in a renegotiation-proof manner does not require exact knowledge of the time preferences.

Recall that the criterion for renegotiation-proofness invoked here requires that the SPE is not Pareto-dominated by any possible outcomes one period hence, but it does not require Pareto efficiency among contemporary outcomes. Note indeed that there are nodes in this implementation game such that the SPE outcome, as evaluated at them, is not Pareto-efficient. For example, in stage 2 at the node that follows a buyer's offer $q < s$, the SPE involves rejection of this offer and hence an inefficient one-period delay. If, instead, we required efficiency at each node, we would be back with the renegotiation-proofness criterion of Section III, and nothing would be gained from the added structure. More precisely, we make the following claim.

Claim: Modify Definition 4 to read: "for all s and b, any node x and any SPE e, the outcome $O(s,b,x,e)$ is Pareto-efficient." Then, the only functions P that are renegotiation-proof implementable over time are the constant ones.

PROOF:

Consider a function P that 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 an 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 of 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 an SPE for $s_{\max}$ and $b_{\min}$.

V. Discussion

A. *Costly Renegotiation*

One may question the sense in which the implementation game of Section IV is renegotiation-proof. This is because there exist

[7]This idea is reminiscent of requirements made in the context of virtual implementation by Hitoshi Matsushima (1988) and Dilip Abreu and Arunva Sen (1991).

subgames at which the SPE is not efficient, so that the parties would have an incentive to renegotiate to an efficient outcome. However, if renegotiation is costly, this game may be immune to such criticism. Specifically, if the renegotiation process is itself time-consuming and 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 IV is renegotiation-proof in the following sense. At any node of the implementation game, it will not be profitable for both parties to renegotiate, since by Definition 4 any alternative agreement that would be implemented after at least a one-period delay 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 may still be made, depending on the form of these costs.

B. *The Renegotiation-Proofness Requirement: Comparison with the Literature*

We have already noted that our approach to renegotiation-proofness is somewhat different from the approach taken in the literature we cite. We speak about the notion of a renegotiation-proof contract (or implementation game), while the literature cited above looks at implementation subject to the constraint that terminal nodes can be renegotiated through some given bargaining process. To think about this difference more concretely, let us return to the world of Section III (before the introduction of the time dimension), where renegotiation-proofness is associated with some costless possibility of renegotiating inefficient outcomes. The work we cited assumes that, for all s, b, and terminal-node y, the outcome of renegotiation between s and b starting from y is known.[8] Let the function $g(s,b,y)$ describe the outcome of such renegotiation. Given g, one may study the full consequences of a contract, even if it is not renegotiation-proof. If the realized valuations are s and b, the parties to such a contract will simply treat terminal-node y as if the outcome $g(s,b,y)$ is actually attached to it.

Notice that, since the bargaining function g is allowed to depend on s and b, there are price functions that are implementable given some function g but are not implementable in a renegotiation-proof manner in the sense of the present paper. For example, given the "split-the-difference" bargaining function $g(s,b,\mathrm{D})=(s+b)/2$, the price function $P(s,b)=(s+b)/2$ is implementable by the trivial contract that specifies just the outcome D. However, as we know from Proposition 1, since this price function is not constant, it cannot be implemented in a renegotiation-proof manner according to our approach.

This observation then raises the question of whether our approach is unduly restrictive; that is, in what sense are price functions implementable by our renegotiation-proof contract the only relevant ones? To address this question, consider a price function P that cannot be implemented by a renegotiation-proof contract in the sense of Section III but can be implemented by a contract augmented by some bargaining function g in the sense explained above. Now, assuming that the bargaining process summarized by the function g can be described as a game in extensive form, it should be possible to complete a non-renegotiation-proof contract by substituting this game for terminal nodes that prescribe the outcome D. Further, if $g(s,b,\mathrm{D})$ is the result of different s and b types engaging in different bargaining games, the contract may be extended to include the game that decides what bargaining games s and b will follow, and so on. Notice that the fact that some moves in these bargaining games may be nonverifiable poses no problem: whatever makes the parties take these moves in

[8] This formulation is more clearly associated with Maskin and Moore (1988) than with Aghion et al. (1989), who take a closer look at the renegotiation process and its design. However, abstracting from the details of the different approaches, both these papers as well, as the work of Chung (1988) and of Green and Laffont (1988) essentially fit into this formulation.

voluntary bargaining will presumably make them take these moves when this bargaining game is played as part of the contract. Now, for the same reasons that inefficient outcomes of the contract will be renegotiated, the bargaining game itself may not involve inefficiencies, since presumably they will be renegotiated too. It follows then that the compounded contract has to be renegotiation-proof in the sense of this paper. Thus, roughly speaking, once we take the logic of the renegotiation idea a few steps further, it seems to imply that the price functions implementable by renegotiation-proof contracts in the sense of this paper are perhaps the only ones that can somehow be implemented in this environment.

The above argument is obviously not complete. It implicitly assumes, for example, that the same solution concept prevails in the contract and the bargaining games used for renegotiation. Thus, while it deserves perhaps more careful consideration before we embrace it, this argument at least suggests that the focus on renegotiation-proof contracts is not arbitrary.

C. *How Renegotiation-Proofness Affects the Consequences of Nonverifiability*

Nonverifiability of information is a form of imperfection. Like 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, to assure proper investment incentives, the contract has to condition the outcomes on the information. However, when the information is nonverifiable, a contract that conditions on it may not be enforceable, and its absence might result in an inefficient investment level.

Obviously, when the possibility of renegotiation exists so that the relevant contracts are the renegotiation-proof ones, the inefficiency/underinvestment problem pointed out above may become more pronounced (see Green and Laffont [1988] for a discussion of this issue). Recall the example presented in the Introduction: $S=\{0\}$; $B=\{1,2\}$; both of the buyer's types have *ex ante* equal probability; to produce the unit, the seller has to invest the sum 1.2 *before* the buyer's valuation is determined. Now, the renegotiation-proof contracts in the sense of Proposition 1 include only the constant-price ones, with price between 0 and 1. Thus, if that is the right renegotiation-proofness restriction, the seller cannot recover his investment and hence will inefficiently avoid production. The implication of Section IV and Proposition 2 for this example is that the inefficiency derived here is due to the too powerful renegotiation-proofness criterion. If renegotiation is time-consuming or otherwise costly to the extent that the criterion of Section IV is appropriate, then the set of relevant contracts is richer, and hence the potential inefficiency problem seems less severe. Here, the contract that implements $P(0,1)=0.9$ and $P(0,2)=1.7$ is renegotiation-proof. This contract alleviates the inefficiency by making the investment of 1.2 profitable.

D. *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. Three properties of time are used here: (i) time is costly; (ii) its passage is irreversible; (iii) the cost of time may be related to the basic valuations. Note that property (iii) is inherent to the time preferences used throughout [i.e., buyers with the same δ but different b's will differently rank (p,t) vs. $(q,t+1)$ for some p and q] and, in fact, to any time preferences that involve some form of discounting.

Of course, the analysis does not necessitate the incorporation of a real time dimension into the contract. The effects of time may be mimicked by imposing other costs on the parties, as long as such costs bear some relationship to the underlying valuations so as to facilitate the separation. 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.

E. *Enforcement Difficulties*

In this, as well as in any other implementation model, the contract (or implementation game) has to be enforceable in two senses. First, the moves prescribed by the contract should be enforceable. Second, it has to be possible to enforce that no *additional* payoff-relevant moves are made. Obviously, if the parties can make such additional moves (say, cause damages to each other), the implementation game and hence its outcome may be quite different from what is specified by the contract. The first of these aspects is addressed by the requirement that the moves prescribed by the contract are verifiable to the authority that oversees the implementation stage (the "court"). The second is usually not addressed explicitly. The implicit assumption in this respect is that any such additional payoff-relevant moves are either of no consequence for the game or are verifiable and can be prevented by the "court."

Both of these aspects of enforcement are obviously more problematic when the implementation procedure is relatively complex. Since the procedure followed in the implementation stage here is both rather elaborate and may last over time, it may naturally be susceptible to this criticism.

F. *A Final Remark*

The basic seller–buyer scenario that we analyze has a rather special structure. Two major characteristics of this scenario are that (i) the parties' preference rankings of all outcomes except the no-sale outcome are diametrically opposed and (ii) all types of seller (or buyer) have almost identical preferences and differ only with respect to how they rank the no-sale outcome vis-à-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 speculation on the theoretical possibilities of contracting with nonverifiable information, which is much beyond the boundaries of the above discussion. 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 the confines of the present model.

REFERENCES

Abreu, Dilip and Sen, Arunva, "Virtual Implementation in Nash Equilibrium," *Econometrica*, July 1991, *59*, 997–1021.

Aghion, Philip, Dewatripont, Mathias and Rey, Patrick, "Renegotiation Design Under Symmetric Information," mimeo, Harvard University, 1989.

Chung, Tai-Young, "Risk Sharing, Specific Investment, and Incomplete Contracts," mimeo, Northwestern University, 1988.

Glazer, Jacob and Ma, Albert, "Efficient Allocation of a Prize—King Solomon's Dilemma," *Games and Economic Behavior*, June 1989, *1*, 222–33.

Grossman, Sanford and Hart, Oliver, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, August 1986, *94*, 691–719.

Green, Jerry and Laffont, Jean-Jacques, "Renegotiation and the Form of Efficient Contracts," mimeo, Harvard University, 1988.

Hart, Oliver and Moore, John, "Incomplete Contracts and Renegotiation," *Econometrica*, July 1988, *56*, 755–86.

Maskin, Eric and Moore, John, "Implementation and Renegotiation," mimeo, Harvard University, 1988.

Matsushima, Hitoshi, "A New Approach to the Implementation Problem," *Journal of Economic Theory*, June 1988, *45*, 128–44.

Moore, John and Repullo, Rafael, "Subgame Perfect Implementation," *Econometrica*, September 1988, *56*, 1191–1220.

Rubinstein, Ariel and Wolinsky, Asher, "Renegotiation-Proof Implementation and Time Preferences," Working Paper 14-90, Tel-Aviv University, 1990.

Tirole, Jean, "Procurement and Renegotiation," *Journal of Political Economy*, April 1986, *94*, 235–59.