

## COSTLY BARGAINING AND RENEGOTIATION

BY LUCA ANDERLINI AND LEONARDO FELLI<sup>1</sup>

We identify the inefficiencies that arise when negotiation between two parties takes place in the presence of transaction costs. First, for some values of these costs it is efficient to reach an agreement but the unique equilibrium outcome is one in which agreement is never reached. Secondly, even when there are equilibria in which an agreement is reached, we find that the model always has an equilibrium in which agreement is never reached, as well as equilibria in which agreement is delayed for an arbitrary length of time.

Finally, the only way in which the parties can reach an agreement in equilibrium is by using inefficient punishments for (some of) the opponent's deviations. We argue that this implies that, when the parties are given the opportunity to renegotiate out of these inefficiencies, the only equilibrium outcome that survives is the one in which agreement is never reached, regardless of the value of the transaction costs.

KEYWORDS: Optional bargaining costs, inefficient bargaining outcomes, renegotiation, imperfect recall.

### 1. INTRODUCTION

#### 1.1. *Motivation*

THE COASE THEOREM (Coase (1960)) is one of the cornerstones of modern economic analysis. It shapes the way economists think about the efficiency or inefficiency of outcomes in most economic situations. It guarantees that, if property rights are fully allocated, economic agents will exhaust any mutual gains from trade. Fully informed rational agents, unless they are exogenously restricted in their bargaining opportunities, will ensure that there are no *unexploited gains from trade*.

This view of the (necessary) exploitation of all possible gains from trade is at the center of modern economic analysis. Economists faced with an inefficient outcome of the negotiation between two rational agents will automatically look for reasons that impede full and frictionless bargaining between the agents.

In this paper we focus on the impact of *transaction costs* on the Coase theorem. We show that, in a complete information world, transaction costs

<sup>1</sup>We are grateful to Vincent Crawford, Joseph Harrington, Roger Lagunoff, Eric Maskin, Matthew Jackson, Steven Matthews, Larry Samuelson, Ilya Segal, a co-editor and two anonymous referees for insightful comments. We would also like to thank seminar participants at Brown, Cambridge, CentER (Tilburg), Ente Einaudi (Rome), Erasmus (Rotterdam), Essex, E.S.S.E.T. 1999 (Gerzensee), Georgetown, John's Hopkins, The London School of Economics, Northwestern, S.E.D. 1999 (Alghero), S.I.T.E. 1998 (Stanford), The Stockholm School of Economics, Tel-Aviv University, and the University of Venice for helpful discussions and feedback. The research work for this paper was carried out while we were visiting the Department of Economics at the University of Pennsylvania. We are grateful for their stimulating hospitality. Of course, we are solely responsible for any remaining errors. We acknowledge financial support from the E.S.R.C. (Grant R000237825).

imply that the Coase theorem no longer holds in the sense that an efficient outcome is no longer guaranteed. In the model that we analyze, for certain values of the transaction costs only inefficient equilibria are possible, while for other values of these costs both efficient and inefficient equilibria obtain. In the latter case we find that it is not possible to select the efficient outcomes in a consistent way: there are no equilibria of the model that guarantee an efficient outcome in every subgame.

Given the impossibility of selecting efficient outcomes by fiat, we proceed as follows. Keeping as given the friction introduced by the transaction costs, we expand the negotiation possibilities for the two agents—we build into the extensive form opportunities for the parties to break out of inefficient outcomes. We find that in this case the *only* equilibrium outcome that survives is the most inefficient possible one: agreement is never reached and the entire surplus fails to materialize.

### 1.2. *Costly Bargaining*

Our point of departure is the leading extensive form model of negotiation between two parties, namely an alternating offers bargaining game with complete information with potentially infinitely many rounds of negotiation in which the players discount the future at a strictly positive rate (Rubinstein (1982)).<sup>2</sup> We introduce transaction costs in the following way. Both parties, at each round of negotiation, must pay a positive cost to participate in that round of the bargaining game. At each round, both parties have a choice of whether or not to pay their respective participation costs. Each round of negotiation takes place only if both parties pay their participation costs. If either player decides not to pay, the negotiation is postponed until the next period.

The interpretation of the participation costs that we favor is the following. At the beginning of each period, both parties must decide (irrevocably for that period) whether to spend that period of time at the negotiation table, or to engage in some other activity that yields a positive payoff. The participation costs in our model can simply be thought of as these alternative payoffs that the agents forego in order to engage in the negotiation activity for that period.

Obviously, the bargaining situations that our model fits best are those in which the participation costs we have described are a prominent feature of the bargaining process. First of all, the alternative payoff that the parties can earn, although smaller, must be of the same order of magnitude as the potential payoff from a bargaining agreement. Secondly, the time necessary to carry out each round of bargaining cannot be negligible. Offers and counteroffers might involve intricate details of the transaction at hand that take time to describe, check, and verify.

<sup>2</sup> Many of our arguments are based on modifications of the proof that the Rubinstein (1982) model has a unique subgame perfect equilibrium presented by Shaked and Sutton (1984).

The first sense in which the Coase theorem fails in our model is the following. There exist values of the participation costs such that it is efficient for the parties to reach an agreement (the sum of the costs is strictly smaller than the size of the surplus) and yet the *unique* equilibrium of the game is for the parties never to pay the costs so that an agreement is never reached (Theorems 1 and 2 below).

Having established Theorems 1 and 2 below, we focus on the case in which the values of the participation costs are low enough so that the parties will be able to reach an agreement in equilibrium. In this case the model displays a wide variety of equilibria: (efficient) equilibria with immediate agreement (Theorem 3 below), (inefficient) equilibria with an agreement with an arbitrarily long delay (Theorem 4 below), and (inefficient) equilibria in which an agreement is never reached (Theorem 1 below). Therefore, the Coase theorem fails in this case too in the sense that it is no longer *necessarily* the case that the outcome of the bargaining between the parties is Pareto efficient.

In the case in which the participation costs are such that there are both efficient and inefficient equilibria, a natural reaction is that it is just a matter of choosing the right selection criterion to be able to isolate the efficient equilibria. If this were possible one would conclude that, in a sense, the Coase theorem does not fail in this setting for low enough transaction costs. In Section 5 below, we show that this way of proceeding does not work in our model. The reason is that all equilibrium agreements are sustained by off-the-equilibrium-path inefficient continuation equilibria needed to punish the players for not paying their participation costs. Since all efficient equilibria must clearly prescribe that an agreement takes place, it follows that a selection criterion that implies efficiency in a consistent way across every subgame does not work in our set-up. In fact, the set of equilibria that survives any such selection criterion is empty in our model (Theorem 6 below).

The fact that inefficient equilibrium outcomes are possible in our model leads naturally to the question of whether the source of the inefficiency and the failure of the Coase theorem lies in the limited negotiation opportunities given to the parties. To address this question we proceed in the following way. We modify the extensive form of the game so as to allow the parties a chance to start a fresh negotiation whenever they are playing strategies that put them strictly within the Pareto frontier of their payoffs. We do this by modifying the extensive form of the game and transforming it into a game of imperfect recall. We assume that, at the beginning of each period, with strictly positive probability, the parties do not recall the past history of play. This affords them a chance to renegotiate out of inefficient punishments. The result is devastating for the equilibria in which agreement is reached. When the probability of forgetting the history of play is above a minimum threshold (smaller than one), the unique equilibrium outcome of the modified game is for the parties never to pay the costs and therefore never to reach an agreement. This is true regardless of the size of the participation costs, provided of course that they are positive.

We view this as the most serious failure of the Coase theorem in our model. If one expands the parties' opportunities to bargain, the inefficiency becomes extreme. Agreement is never reached, whatever the size of the transaction costs.

### 1.3. *Related Literature*

As we mentioned above, the inefficiency results that we obtain in this paper can be viewed as a failure of the Coase theorem in the presence of transaction costs.<sup>3</sup>

It is clear that the original version of the Coase theorem (Coase (1960)) explicitly assumes the absence of any transaction costs or other frictions in the bargaining process. Indeed, Coase (1992) describes the theorem as a provocative result that was meant to show how unrealistic is the world without transaction costs.<sup>4</sup> It should, however, be noticed that, sometimes, subsequent interpretations of the original claim have strengthened it way beyond the realm of frictionless negotiation.<sup>5</sup> It does not seem uncommon for standard microeconomics undergraduate texts to suggest that the Coase theorem should hold in the presence of transaction costs.<sup>6</sup>

The analysis in Anderlini and Felli (1997) is also concerned with the hold-up problem generated by ex-ante contractual costs in a stylized contracting model and with the inefficiencies it generates. However, the main concern in Anderlini and Felli (1997) is with the robustness of the inefficiencies to changes in a number of assumptions. In particular, that paper focuses on the nature of the costs payable by the parties to make the contracting stage feasible, and on the possibility that the parties may rely on an expanded contract that includes contracting on the ex-ante costs themselves. By converse, in this paper we take it as given that the parties bargain according to a given protocol, and that they have to pay their participation costs in order to negotiate at each round.

<sup>3</sup> We are certainly not the first to point out that the Coase theorem no longer holds when there are frictions in the bargaining process. There is a vast literature on bargaining models where the frictions take the form of incomplete and asymmetric information. With incomplete information, efficient agreements often cannot be reached and delays in bargaining may obtain. (See Muthoo (1999) for an up-to-date coverage as well as extensive references on this strand of literature and other issues in bargaining theory.) By contrast, the main bargaining game that we analyze here is one of complete information. The source of inefficiencies in this paper can therefore be traced directly to the presence of participation costs.

<sup>4</sup> de Meza (1988) provides an extensive survey of the literature on the Coase theorem, including an outline of its history and possible interpretations.

<sup>5</sup> By contrast, Dixit and Olson (1997) have recently been concerned with a classical Coasian public good problem in which they explicitly model the agents' ex-ante (possibly costly) decisions of whether to participate or not in the bargaining process. In this context, they find both efficient and inefficient equilibria. They also highlight the inefficiency of the symmetric (mixed-strategy) equilibria of their model.

<sup>6</sup> For instance, an excellent textbook widely in use in the U.S. and elsewhere claims that, in its strongest formulation, the Coase theorem is interpreted as guaranteeing an efficient outcome whenever the potential mutual gains "exceed [the] necessary bargaining costs" (Nicholson (1989, p. 726)).

Because they are sunk by the time offers are made and accepted or rejected, the participation costs that we introduce in the bargaining problem generate a version of the hold-up problem. This is the main source of inefficiency in the models that we analyze in this paper.

The need for relationship-specific investment may allow one party to hold-up the other when fully contingent contracts are not available (Klein, Crawford, and Alchian (1978), Grout (1984), Williamson (1985), Grossman and Hart (1986)). This key observation has generated a large and varied literature that has shed light on many central issues ranging from vertical and lateral integration (Grossman and Hart (1986)), ownership rights (Hart and Moore (1990)), the delegation of authority (Aghion and Tirole (1997)) and power (Rajan and Zingales (1998)) within firms. In all these models, a hold-up problem arises because the only possible contracts are incomplete. In a sense this causality is reversed in this paper. Here, the hold-up problem generated by the participation costs may induce inefficient bargaining outcomes; in some cases it may prevent the parties from reaching an agreement at all. The lack of agreement in a bargaining problem, in turn, can be viewed as an extreme form of contractual incompleteness. In a way, it is the hold-up problem generated by the participation costs that is the cause of contractual incompleteness rather than vice-versa: the parties do not sign a contract when in fact it would be efficient to do so.

A small number of recent papers has been concerned with inefficiencies that might arise in bargaining models with complete information. The extensive form games, and hence the sources of inefficiencies, that they analyze are substantially different from ours. In Fernandez and Glazer (1991) and Busch and Wen (1995) the nature of the bargaining costs is the exact opposite to the one tackled here. The parties may choose to pay a cost to delay the negotiation for a period. They both find efficient and inefficient equilibria in their models. Fershtman and Seidmann (1993) analyze a bargaining model in which inefficient equilibria arise because of the nonstationarity of the game. The nonstationarity of their game is given by the presence of a deadline and by the fact that each party cannot accept an offer that he has rejected in the past. Riedl (1997) analyzes a model in which only one player incurs a cost to participate in the bargaining process. He concentrates on the comparison of the case in which the cost is payable once with the case in which a cost is payable in each period.

#### 1.4. *Overview*

The paper is organized as follows. In Section 2 we describe in detail our model of alternating offers bargaining with transaction costs. Section 3 contains our first inefficiency result and a characterization of the equilibria of the model described in Section 2. In Section 4 we investigate the robustness of the inefficient and of the efficient equilibria of our model to some basic changes in the description of the game. In Section 5 we show that it is impossible to select the Pareto efficient equilibria of our game in a way that is consistent across subgames. Section 6 contains our model of renegotiation opportunities in the

extensive form. Here, we present our second main result—namely the fact that the only equilibrium outcome of our game of imperfect recall is that agreement is never reached. Section 7 briefly concludes the paper. For ease of exposition all proofs are relegated to the Appendix.

## 2. THE MODEL

We consider a bargaining game between two players indexed by  $i \in \{A, B\}$ . The game consists of potentially infinitely many rounds of alternating offers  $n = 1, 2, \dots$  and the size of the surplus to be split between the players is normalized to one. Each player  $i$  has to pay a participation cost at round  $n$  denoted  $c_i$  (constant through time). We interpret this cost as the opportunity cost to player  $i$  of the time the player has to spend in the next round of bargaining.<sup>7</sup> Throughout the paper, we focus on the case in which  $c_A + c_B \leq 1$ .

In all odd periods,  $n = 1, 3, 5, \dots$ , player  $A$  has the chance to make offers, and player  $B$  the chance to respond. In all even periods  $n = 2, 4, 6, \dots$ , the players' roles are reversed,  $B$  is the proposer, while  $A$  is the responder. We refer to the odd periods as *A periods* and to even periods as *B periods*.

The size of the surplus to be split between the players is normalized to one. Any offer made in period  $n$  is denoted by  $x \in [0, 1]$ . This denotes  $A$ 's share of the pie, if agreement is reached in period  $n$ . The discount factor of player  $i \in \{A, B\}$  is denoted by  $\delta_i \in [0, 1]$ .

To clarify the structure of each round of bargaining, it is convenient to divide each time period in three *stages*. In stage I of period  $n$ , both players decide simultaneously and independently, whether to pay the costs  $c_i$ . If both players pay their participation costs, then the game moves to stage II of period  $n$ . At the end of stage I, both players observe whether or not the other player has paid his participation cost. If one, or both, players do not pay their cost, then the game moves directly to stage I of period  $n + 1$ .

In stage II of period  $n$ , if  $n$  is odd,  $A$  makes an offer  $x \in [0, 1]$  to  $B$ , that  $B$  observes immediately after it is made. At the end of stage II of period  $n$ , the game moves automatically to stage III of period  $n$ . If  $n$  is even, the players' roles in stage II are reversed.

In stage III of period  $n$ , if  $n$  is odd,  $B$  decides whether to accept or reject  $A$ 's offer. If  $B$  accepts, the game terminates, and the players receive the payoffs described in (1) below. If  $B$  rejects  $A$ 's offer, then the game moves to stage I of period  $n + 1$ . If  $n$  is even, the players' roles in stage III are reversed.

<sup>7</sup> Of course, it is possible that as the parties progress into further rounds of bargaining, they may become more efficient in their use of time. Depending on the particular bargaining situation at hand, offers and counteroffers may become routine, and the time needed for each round of bargaining may shrink. Clearly in this case, the participation costs would be decreasing rather than constant through time. Our first inefficiency result below (the only if part of Theorem 2) applies unchanged if we consider the lower bounds (over time) of any time-dependent participation costs.

The players' payoffs consist of their shares of the pie (zero if agreement is never reached), minus any costs paid, appropriately discounted. To describe the payoffs formally, it is convenient to introduce some further notation at this point. Let  $(\sigma_A, \sigma_B)$  be a pair of strategies for the two players in the game we have just described, and consider the *outcome path*  $\mathcal{O}(\sigma_A, \sigma_B)$  that these strategies induce.<sup>8</sup> Let also  $\mathcal{E}_i(\sigma_A, \sigma_B)$  be the *total* of participation costs that player  $i$  pays along the entire outcome path  $\mathcal{O}(\sigma_A, \sigma_B)$ , discounted at the appropriate rate.

If the outcome path  $\mathcal{O}(\sigma_A, \sigma_B)$  prescribes that the players agree on an offer  $x$  in period  $n$ , then the payoffs to  $A$  and  $B$  are respectively given by

$$(1) \quad \begin{aligned} \Pi_A(\sigma_A, \sigma_B) &= \delta_A^n x - \mathcal{E}_A(\sigma_A, \sigma_B) \quad \text{and} \\ \Pi_B(\sigma_A, \sigma_B) &= \delta_B^n (1 - x) - \mathcal{E}_B(\sigma_A, \sigma_B) \end{aligned}$$

while if the outcome path  $\mathcal{O}(\sigma_A, \sigma_B)$  prescribes that the players never agree on an offer, then the payoff to player  $i \in \{A, B\}$  is given by

$$\Pi_i(\sigma_A, \sigma_B) = -\mathcal{E}_i(\sigma_A, \sigma_B).$$

### 3. SUBGAME PERFECT EQUILIBRIA

In this section we provide a full characterization of the set of subgame perfect equilibria of the alternating offer bargaining game described in Section 2 above.

We first show that the bargaining game always has a subgame perfect equilibrium (henceforth SPE) in which the players do not ever pay the costs and hence agreement is never reached. By construction, this can be proved considering the following pair of strategies that constitute an SPE of the game. Both players do not pay their participation costs in stage I of any period, regardless of the previous history of play. In stage II of any period (off the equilibrium path) the proposing player demands the entire pie for himself. In stage III of any period (again off the equilibrium path) the responding player accepts any offer  $x \in [0, 1]$ . Thus, we have proved our first result.

**THEOREM 1:** *Consider the alternating offers bargaining game with participation costs described in Section 2. Whatever the values of  $\delta_i$  and  $c_i$  for  $i \in \{A, B\}$ , there exists an SPE of the game in which neither player pays his participation cost in any period, and therefore an agreement is never reached.*

We now proceed to characterize the necessary and sufficient conditions on the pair of costs  $(c_A, c_B)$  and the parties' discount factors  $(\delta_A, \delta_B)$  under which the parties are able to achieve an agreement.

<sup>8</sup> Throughout the paper, we focus on pure strategies only. This greatly simplifies the analysis and dramatically reduces the amount of notation we need. The nature of our results would be unaffected by considering equilibria in which players are allowed independent randomizations (behavioral strategies). In particular, the analogues of Theorems 2 and 5 below hold when mixing (behavioral strategies) is allowed.

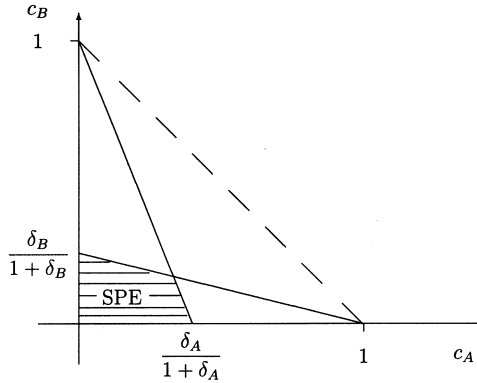


FIGURE 1.—SPE with agreement in finite time.

**THEOREM 2:** Consider the alternating offers bargaining game with participation costs described in Section 2. The game has an SPE in which an agreement is reached in finite time if and only if  $\delta_i$  and  $c_i$  for  $i \in \{A, B\}$  satisfy

$$(2) \quad \delta_A(1 - c_A - c_B) \geq c_A \quad \text{and} \quad \delta_B(1 - c_A - c_B) \geq c_B.$$

For given  $\delta_A$  and  $\delta_B$ , the set of costs  $(c_A, c_B)$  for which an agreement is reached is represented by the shaded region in Figure 1.

A complete proof of Theorem 2 appears in the Appendix. It is useful to outline here the steps of the argument that proves that the inequalities in (2) are necessary for the existence of an SPE with agreement in finite time.

Assume that an SPE with an agreement in finite time exists. Clearly, in any SPE the equilibrium agreement must satisfy  $x \in [c_A, 1 - c_B]$ . This is because the parties' payoffs cannot be negative in any SPE.<sup>9</sup> From the stationarity of the game it follows that if an SPE with agreement in finite time exists, then there must be some SPE with immediate agreement in every subgame starting in stage I of every period.

Consider now stage III of a period in which the costs have been paid and  $B$  has made an offer to  $A$ . Clearly  $A$  will accept all offers  $x$  that are above  $\delta_A(x_A^H - c_A)$ , where  $x_A^H$  is the highest possible equilibrium agreement in a period in which  $A$  is the proposer. Using subgame perfection we can now conclude that the highest possible equilibrium agreement in a period in which  $B$  is the proposer,  $x_B^H$ , satisfies

$$(3) \quad x_B^H \leq \delta_A(x_A^H - c_A).$$

A completely symmetric argument proves that

$$(4) \quad 1 - x_A^L \leq \delta_B(1 - x_B^L - c_B)$$

<sup>9</sup>This is immediate from the fact that each player can guarantee a payoff of zero by never paying his participation cost.



where  $x_A^L$  (respectively  $x_B^L$ ) is the lowest possible equilibrium agreement in a period in which  $A$  (respectively  $B$ ) is the proposer.

Recall now that all equilibrium agreements must be in the range  $[c_A, 1 - c_B]$ . Therefore, we can now substitute  $x_B^H \geq c_A$ ,  $x_A^H \leq 1 - c_B$ ,  $x_A^L \leq 1 - c_B$ , and  $x_B^L \geq c_A$  into (3) and (4) to obtain the inequalities in (2), and hence conclude the argument.

Clearly, Theorem 2 supports our first inefficiency claim. The sum of the participation costs is less than the total available surplus anywhere below the dashed line in Figure 1. Given any pair of discount factors, there exists a region of possible participation costs such that the model has a unique, inefficient, SPE outcome. In Figure 1, for any pair  $(c_A, c_B)$  below the dashed line but outside the shaded area, the participation costs add up to less than one, but no agreement is ever reached.

We are now ready to give a more detailed characterization of the SPE with agreements of this game. We start by identifying the range of possible equilibrium shares of the pie in every possible subgame when agreement is immediate.

**THEOREM 3:** *Consider the alternating offers bargaining game with participation costs described in Section 2, and assume that  $\delta_i$  and  $c_i$  for  $i \in \{A, B\}$  are such that (2) holds so that the game has some SPE in which an agreement is reached in finite time. Consider any subgame starting in stage I of any  $A$  period (the  $A$  subgames from now on).<sup>10</sup> Then there exists an SPE of the  $A$  subgames in which  $x_A$  is agreed immediately, if and only if*

$$(5) \quad x_A \in [1 - \delta_B(1 - c_A - c_B), 1 - c_B].$$

*Symmetrically, consider any subgame starting in stage I of any  $B$  period (the  $B$  subgames from now on). Then there exists an SPE of the  $B$  subgames in which  $x_B$  is agreed immediately, if and only if*

$$(6) \quad x_B \in [c_A, \delta_A(1 - c_A - c_B)].$$

Our next result both completes our characterization of the set of SPE payoffs, and supports our second inefficiency claim. Every sharing of the pie that can be supported as an immediate agreement can also take place with a delay of an arbitrary number of periods.

**THEOREM 4:** *Consider the alternating offers bargaining game with participation costs described in Section 2, and assume that  $\delta_i$  and  $c_i$  for  $i \in \{A, B\}$  are such that (2) holds so that the game has some SPE in which an agreement is reached in finite time.*

*Let any  $x_A$  as in (5) and any odd number  $n$  be given. Then there exists an SPE of the  $A$  subgames in which the (continuation) payoffs to the players are respectively given by*

$$(7) \quad \Pi_A = \delta_A^n(x_A - c_A), \quad \Pi_B = \delta_B^n(1 - x_A - c_B).$$

<sup>10</sup> Recall that we refer to all odd periods as  $A$  periods, and to all even periods as  $B$  periods.

Moreover, let any  $x_B$  as in (6) and any even number  $n$  be given. Then there exists an SPE of the  $A$  subgames in which the (continuation) payoffs to the players are respectively given by

$$(8) \quad \Pi_A = \delta_A^n(x_B - c_A), \quad \Pi_B = \delta_B^n(1 - x_B - c_B).$$

Symmetrically, let any  $x_B$  as in (6) and any odd number  $n$  be given. Then there exists an SPE of the  $B$  subgames in which the (continuation) payoffs to the players are as in (8). Moreover, let any  $x_A$  as in (5) and any even number  $n$  be given. Then there exists an SPE of the  $B$  subgames in which the (continuation) payoffs to the players are as in (7).

#### 4. ROBUSTNESS OF EQUILIBRIA

In this section, we carry out four robustness exercises about the SPE of the game described in Section 2 that we have identified in Section 3.

Our first concern is the relationship between the set of SPE of our game with the set of SPE of a finite version of the same game (Ståhl (1972)). The unique SPE identified by Rubinstein (1982) of the same bargaining game when there are no participation costs has many reassuring properties. Among these is the fact that if a version of the same game with a truncated time horizon is considered, the limit of the SPE of the finite games coincides with the unique SPE of the infinite horizon game. This is not the case in our bargaining model with participation costs. In fact when we truncate the time horizon to be finite in our model, the only possible SPE outcome is the one in which neither player ever pays his participation cost and hence no agreement is reached provided only that participation costs are positive.

The intuition behind Remark 1 below is a familiar backward induction argument. No agreement is possible in the last period since the responder would have to get a share of zero if agreement is reached, and therefore he will not pay his participation cost in that period. This easily implies that no agreement is possible in the last period but one, and so on.

Let  $\Gamma^\infty$  represent the infinite horizon alternating offers bargaining game with participation costs described in Section 2. For any finite  $N > 1$ , let  $\Gamma^N$  represent the same game with time horizon truncated at  $N$ . In other words, in  $\Gamma^N$ , if period  $N$  is ever reached, the game terminates, regardless of whether an agreement has been reached or not. If no agreement has been reached by period  $N$ , the players' payoffs are zero, minus any costs paid of course. We can then state the following.

REMARK 1: *Let any finite  $N \geq 1$  be given. Then the unique SPE outcome of  $\Gamma^N$  is that neither player pays his participation cost in any period and hence agreement is never reached.*

Trivially, Remark 1 implies that the only SPE outcome of  $\Gamma^\infty$  that is in fact the limit of any sequence of SPE outcomes of  $\Gamma^N$  as  $N$  grows is the one in which agreement is never reached.

Our next concern is the robustness of the SPE in which neither party ever pays his participation cost and hence no agreement is ever reached to the sequential payments of the participation costs. It is a legitimate concern to check whether this equilibrium is attributable to a simple coordination failure or whether it depends on other features of the structure of our alternating offers bargaining game with participation costs. It turns out that this SPE is indeed robust to the players paying their participation costs sequentially, before any offer is made and accepted or rejected. We therefore conclude that, while a coordination failure is clearly possible in the game we analyze, it is not the ultimate source of inefficiencies in our set-up.

Let  $\mathcal{S}$  be any sequence of the form  $\{i_1, i_2, \dots, i_n, \dots\}$ , where  $i_n \in \{A, B\}$  for every  $n$ . Let  $\Gamma(\mathcal{S})$  be the game derived from the one described in Section 2, modified as follows. In stage I of period  $n$ , player  $i_n$  first decides whether to pay his participation cost or not. Following  $i_n$ 's decision, the other player observes whether  $i_n$  has paid his cost or not, and then decides whether to pay his own participation cost. The description of stages II and III of every period in  $\Gamma(\mathcal{S})$  is exactly the same as for the original game described in Section 2. We are then able to state the following.

REMARK 2: *Fix any arbitrary sequence  $\mathcal{S}$  as described above. Then  $\Gamma(\mathcal{S})$  always has an SPE in which neither player ever pays his participation cost, and hence agreement is never reached.*<sup>11</sup>

Our third robustness exercise concerns the viability of our first inefficiency result when the identity of the proposer does not necessarily follow a strict alternating offers protocol. It turns out that Theorem 2 is in fact robust to a number of changes in the alternating offers nature of the game—it applies to all of the bargaining games with participation costs that we describe below.<sup>12</sup>

Consider the following class of games. For want of a better name we refer to them as bargaining games with participation costs and *weakly* alternating offers. Every period is still divided into three stages as before, the only change from the game described in Section 2 is what determines the identity of the proposer.

In stage II of period 1,  $A$  is the proposer. In all subsequent periods, the identity of the proposer depends on the previous history of play in a general (deterministic) way. However, we impose two restrictions on how the history of play determines who makes an offer at each point.

<sup>11</sup> Notice that this type of equilibrium is always present, even when  $c_A = c_B = 0$ , both when costs are paid simultaneously, and when they are paid sequentially. However, this pure coordination failure disappears if we are willing to eliminate weakly dominated strategies. In both cases, when  $c_A = c_B = 0$ , not paying the participation cost is a weakly dominated strategy for both players. This is, of course, not true when both costs are positive.

<sup>12</sup> Our description is informal to economize on new notation and space.

The first restriction is that of *stationarity*. All subgames starting in stage I of any period in which (if costs are paid of course) player  $i \in \{A, B\}$  is the proposer in stage II are identical.

The second restriction that we impose is that offers must *alternate after a rejection*. Suppose that in stage II of any period, player  $i \in \{A, B\}$  makes an offer to play  $j \neq i$ , and that player  $j$  rejects the offer in stage III of the same period. Then, the next offer, in stage II of any subsequent period in which both players pay their participation costs, is made by player  $j$ .

As we anticipated, Theorem 2 applies to any game in the class we have just described.<sup>13</sup>

**REMARK 3:** *Let any bargaining game with participation costs and weakly alternating offers be given. Then the game has an SPE in which an agreement is reached in finite time if and only if  $\delta_i$  and  $c_i$  for  $i \in \{A, B\}$  satisfy the inequalities in condition (2) of Theorem 2 above.*

Our last concern is also with the robustness of our first inefficiency result to changes in the alternating offers nature of the game. In particular, we now ask whether the inefficiency in Theorem 2 above survives if the identity of the proposer is randomly determined *after* participation costs have been paid.

We consider the simplest modification of the alternating offer bargaining game with participation costs of Section 2 that allows for the identity of the proposer to be randomly determined. We assume that in odd periods  $A$  makes an offer with probability  $p$  and  $B$  makes an offer with probability  $(1 - p)$ , while in even periods the identity of the two players is reversed;  $B$  becomes the proposer with probability  $p$  while it is  $A$  who makes an offer with probability  $(1 - p)$ . Without loss of generality (up to a relabeling of players), we assume that  $p \geq 1/2$ .

This way of introducing a randomly determined proposer seems to be the simplest one that allow us, for different values of  $p$ , to span the whole spectrum of possible random choices of the proposer. For  $p = 1$  the game coincides with the alternating offers bargaining game analyzed above, while for  $p = 1/2$  the players have equal probabilities of making an offer in each period, the game is fully symmetric and player  $A$ 's first mover advantage disappears. Given the random choice of proposer we have just outlined, the extensive form game we want to analyze can be briefly described as follows.

<sup>13</sup> Notice that the class of bargaining games with participation costs and weakly alternating offers is relatively broad. For instance it encompasses games in which the proposer changes or does not change according to whether costs have been paid and by whom. For example, we could postulate that if we are, say, in an  $A$  (respectively  $B$ ) period and  $B$  (respectively  $A$ ) does not pay his participation cost, then the proposer does not change, while  $B$  (respectively  $A$ ) becomes the proposer if he pays his participation cost. Somewhat surprisingly, Remark 3 shows that, for this game, the configurations of parameters such that an SPE with agreement in finite time exists are exactly the same as for the game with strictly alternating offers analyzed in Section 2 above.

Stage I of every period  $n$  is unchanged from the extensive form described in Section 2. At the beginning of stage II of period  $n$  both players observe the realization of a public randomization device that has two possible outcomes  $\alpha$  and  $\beta$ . If the realization is  $\alpha$  then player  $A$  makes an offer  $x \in [0, 1]$  to  $B$  while, if the realization is  $\beta$ , it is  $B$ 's turn to make an offer  $x \in [0, 1]$  to  $A$ . If  $n$  is odd the randomization device draws  $\alpha$  with probability  $p$  and  $\beta$  with probability  $(1 - p)$ . If instead  $n$  is even, the randomization device draws  $\alpha$  with probability  $(1 - p)$  and  $\beta$  with probability  $p$ .

In essence, stage III is also unchanged. The player who has received an offer (in stage II, from the randomly chosen proposer), observes the offer and then decides whether to accept it or reject it. If the offer is accepted the game terminates and the players receive the payoffs described in (1) above. If instead the offer is rejected, the game moves to stage I of period  $n + 1$ .

We can now state the equivalent of Theorem 2 above for the bargaining game we have just described.

**THEOREM 5:** *Consider the bargaining game with random proposer and participation costs described above. The game has an SPE in which an agreement is reached in finite time with positive probability if and only if parameters  $p$ ,  $\delta_i$  and  $c_i$ , for  $i \in \{A, B\}$ , are such that at least one of the following three sets of conditions is satisfied:<sup>14</sup>*

*Condition 1:*

$$(9) \quad p\delta_B(\min\{1 - c_A, p\} - c_B) \geq c_B - (1 - p) \quad \text{and} \\ p\delta_A(\min\{1 - c_B, p\} - c_A) \geq c_A - (1 - p).$$

*Condition 2:*

$$(10) \quad c_A \leq p \quad \text{and} \quad c_B \leq (1 - p).$$

*Condition 3:*

$$(11) \quad c_A \leq (1 - p) \quad \text{and} \quad c_B \leq p.$$

For given  $\delta_A$ ,  $\delta_B$ , and  $p$ , the set of pairs of participation costs  $(c_A, c_B)$  for which an agreement is reached in this new bargaining game is represented by the shaded region in Figure 2.

Using Figure 2 it is immediate to see that the type of inefficiency that we characterized in Theorem 2 above is also present when the identity of the proposer is randomized. Indeed given any pair of discount factors and any probability  $p \in [1/2, 0]$ , there exists a region of possible participation costs such that the model has a unique, inefficient, SPE outcome. In Figure 2, for any pair

<sup>14</sup> Recall that we are focusing on pure strategies throughout. However, it is possible in principle that agreement may be reached in some period conditionally on, say,  $A$  being the proposer, while no agreement is reached if  $B$  becomes the proposer. In this case the probability that an agreement is reached is positive but strictly below one. (See also Lemma A.6 below.)

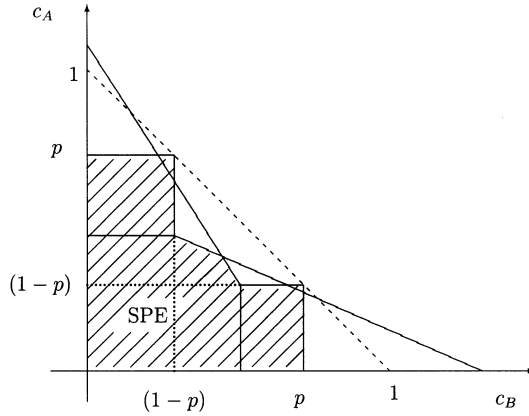


FIGURE 2.—SPE with agreement in finite time and random proposer.

$(c_A, c_B)$  below the dashed line but outside the shaded area, the participation costs add up to less than one, but no agreement is ever reached.

It is also immediate to verify that the conditions in the statement of Theorem 5 are in fact identical to (2) of Theorem 2 when  $p = 1$ . Since the three conditions in the statement of Theorem 5 are continuous in the parameters, Theorem 5 tells us that the inefficiency that we identified in Theorem 2 is robust to small changes in the protocol for choosing the proposer. As  $p$  approaches 1, the set of participation costs for which an agreement can be reached in finite time in the model with random proposer tends to the set of participation costs that yield an agreement in finite time in the game with deterministic alternating offers.

Three further observations about Theorem 5 are in order at this point. Notice first of all that the three conditions in Theorem 5 are not mutually exclusive. In fact, whatever the values of  $p$  and  $\delta_i$  for  $i \in \{A, B\}$ , there is always a region of pairs of participation costs  $(c_A, c_B)$  such that all three conditions are satisfied.

Secondly, the inefficiency identified by Theorem 5 above does not depend on the fact that in each period one player is more likely to make an offer than the other. As Figure 3 shows, Theorem 5 yields an inefficiency region even when  $p = 1/2$  and the game is symmetric in the sense that in every period both players have an equal chance of becoming the proposer.

Finally, agreement cannot always be reached immediately for all values of the costs  $(c_A, c_B)$  for which an SPE with agreement exists (the shaded region in Figure 2). In particular the following remark shows that when Condition 3 in Theorem 5 is satisfied while Condition 1 is *not* satisfied, there does not exist an SPE of the game in which agreement is reached immediately.

REMARK 4: *The bargaining game with random proposer and participation costs has an SPE in which an agreement is reached in period 2 with probability one, but no SPE in which an agreement is reached in period 1 with positive probability if and*

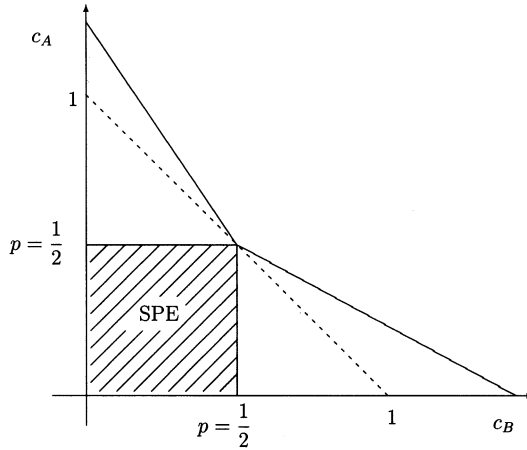


FIGURE 3.—SPE with agreement in finite time, random proposer,  $p = \frac{1}{2}$ .

only if the parameters  $p$ ,  $c_i$  and  $\delta_i$ , for  $i \in \{A, B\}$ , are such that

$$(12) \quad \begin{aligned} c_A &\leq (1 - p) & \text{and} & & c_B &\leq p & \text{and} \\ p\delta_B(p - c_B) &< c_B - (1 - p). \end{aligned}$$

5. CONSISTENTLY PARETO EFFICIENT EQUILIBRIA?

Theorems 3 and 4 tightly characterize the SPE payoffs of the alternating offers bargaining game described in Section 2, when the players agree in finite time on how to divide the available surplus.

On the other hand, Theorem 1 tells us that the game always also has an SPE in which no agreement is reached in finite time. In this SPE, neither player ever pays his participation cost and the players' payoffs are zero.

Thus all the subgames have both Pareto efficient equilibria, in which an agreement is reached immediately (see Theorem 3), and a highly inefficient one in which the surplus is completely dissipated through an infinite delay (see Theorem 1). There are also SPE in which part of the surplus is dissipated since agreement takes place, but is delayed by a finite number of periods (see Theorem 4).

A natural question to ask at this point, and one that is central to this paper, is whether the inefficient SPE of the alternating offers bargaining game with participation costs described in Section 2 can be ruled out.

It is tempting to argue as follows. Since the game at hand is one of complete information, there are no possible strategic reasons for either player to delay agreement. Neither player can possibly hope to accumulate a reputation that will help in subsequent stages of the game. Neither player can possibly gain information about the other player as play unfolds. Therefore, the players will somehow agree to play an efficient equilibrium in which no delays occur. The players will in some way renegotiate out of inefficient equilibria.

This line of reasoning, in our view, is flawed on at least two accounts. The first concerns the modeling of renegotiation in a bargaining game. The second is that, in the game described in Section 2, once renegotiation possibilities are explicitly taken into account, the only SPE that survives is in fact the one in which an agreement is never reached. Therefore the SPE characterized by the most extreme form of inefficiency is the one that is robust to the introduction of renegotiation. Section 6 is entirely devoted to this claim.

The difficulty in taking into account renegotiation possibilities in a bargaining game stems from a simple observation. A bargaining game is, by definition, a model of how the negotiation proceeds between the two players. When they are explicitly modeled, clearly there should be no intrinsic difference between negotiation and renegotiation. Renegotiation is just another round of negotiation, that takes place (ex-post) if the original negotiation has failed to produce an efficient outcome. In short, in a model of negotiation, renegotiation possibilities should be explicitly taken into account in the *extensive form*, rather than grafted as a black box onto the original model. This is what we do in Section 6 below.

In the remainder of this Section, we point out that a simple-minded black box view of renegotiation does not work in the game described in Section 2.

Suppose that, in a Coasian fashion we attempt simply to select for efficient outcomes in our bargaining game with participation costs. A minimal consistency requirement for this operation is that we should recognize that each period of the bargaining game at hand is in fact an entire negotiation game by itself. Therefore, if we believe that efficient outcomes should be selected simply on the grounds that they are efficient, we should now be looking for an SPE that yields an efficient outcome in *every subgame* of the bargaining game. It turns out that this is impossible.

We first proceed with the formal definition of a consistently Pareto efficient SPE and with our next result, and then elaborate on the intuition behind it.<sup>15</sup>

**DEFINITION 1:** An SPE  $(\sigma_A, \sigma_B)$  is called *Consistently Pareto Efficient* (henceforth *CPESPE*) if and only if it yields a Pareto-efficient outcome in every possible subgame.<sup>16</sup>

<sup>15</sup> Various notions of renegotiation-proofness were developed by Benoît and Krishna (1993) (for finitely repeated games), and by Bernheim and Ray (1989), Farrell and Maskin (1989), Farrell and Maskin (1987), and Abreu, Pearce, and Stacchetti (1993) (among others) for infinitely repeated games. Our bargaining game with participation costs, of course is neither a finite game nor a repeated game.

<sup>16</sup> Definition 1 requires efficiency in every possible subgame. From the proof of Theorem 6 it is evident that a weaker definition of CPESPE would suffice. In fact it would be enough to require that a CPESPE yields a Pareto efficient outcome in a subset of subgames—namely every subgame starting at the beginning of every period. We adopt this definition of a CPESPE simply because it seems cleaner in a game-theoretic sense.



We now show that it is impossible to single out an SPE that is consistently Pareto efficient in the way we have just described.

**THEOREM 6:** *Consider the alternating offers bargaining game with participation costs described in Section 2. The set of CPESPE for this game is empty.*

The intuition behind Theorem 6 can be outlined as follows. A CPESPE must yield an agreement in every period, regardless of the history of play that leads to that subgame.<sup>17</sup> Recall that, except for the participation costs, our bargaining game is the original alternating offers bargaining game analyzed by Rubinstein (1982). Once we impose that an agreement must be reached in every period, we can reason about our model in a way that closely parallels well known arguments that apply to the model with no participation costs.

Adapting the argument used by Shaked and Sutton (1984) we can then show the following. First of all if an SPE were to exist with agreement in every period, then there would be a *unique* share of the pie  $x_A$  that is sustainable in equilibrium in every  $A$  subgame, and a unique share of the pie  $x_B$  for every  $B$  subgame. Moreover,  $x_A$  and  $x_B$  have the following property. In stage III of every  $A$  subgame,  $B$  is exactly indifferent between accepting  $A$ 's offer  $x_A$  and rejecting it, and, symmetrically, in stage III of every  $B$  subgame  $A$  is exactly indifferent between accepting  $x_B$  and rejecting it. Therefore, in stage I of every  $A$  period,  $B$  has an incentive not to pay his participation cost: by moving to the next period he earns a payoff that is larger by precisely  $c_B$ . Similarly in stage I of every  $B$  period, player  $A$  can gain  $c_A$  by not paying his cost and forcing the game to move to the next stage.

Theorem 6 implies that to sustain an agreement as an SPE outcome, *inefficient* punishments (off-the-equilibrium-path) are *necessary*. Clearly these must take the form of (off-the-equilibrium-path) delays of one period or more. Definition 1 above is designed to highlight this feature of any SPE involving an agreement in our model.

However, as we stated above, we do not believe that grafting a renegotiation refinement onto a negotiation game is the correct way to proceed. We take Theorem 6 above simply to say that there is no way consistently to select efficient outcomes in our game. Its value lies mainly in clarifying that this is not possible, and in making explicit the sunk cost nature of the intuition behind this fact.

On the basis of Theorem 6 the inefficient SPE of our game have to be granted equal dignity with the efficient ones at this stage of the analysis. In the next section, we proceed to incorporate renegotiation possibilities into the extensive form of the game, and to argue that in this case the SPE with no agreement in finite time is selected among the many possible ones.

<sup>17</sup> Notice that the definition of CPESPE does *not* imply that the *same* agreement must be reached irrespective of history. It only implies that some agreement must be reached in every period, whatever the history of play that leads the players to arrive at the subgame.

## 6. EXTENSIVE FORM RENEGOTIATION

6.1. *Modeling Renegotiation*

In this section we modify the bargaining game described in Section 2 in a way that, in our view, embeds into the extensive form the *chance* for the players to renegotiate out of inefficient outcomes.

We do this in a way that is designed to satisfy three, in our view critical, criteria. First of all, whenever the players find themselves trapped in an inefficient (punishment) phase of play, the extensive form has to give them at least a chance to break out of this inefficient outcome path. Secondly, the possibility of renegotiation must be built into the extensive form as a possibility, rather than an obligation to start afresh and switch to an efficient equilibrium. This is because we want to ensure that our way of tackling the problem here is distinct from the black box renegotiation discussed in Section 5 above. If the extensive form in some way forced efficient play whenever an inefficient outcome path has started, there would be little difference between extensive form renegotiation and black box renegotiation. Our third criterion is closely related to the second one—the extensive form we study must be nontrivial in the sense that it must *allow* in principle for the outcome path both on- and off-the-equilibrium-path to be inefficient. If this were not the case, besides violating our second criterion, via Theorem 6, we would automatically know that the equilibria of the modified extensive form have little to do with the SPE of the original game. This is simply because Theorem 6 tells us that there are no SPE of the original game that yield a Pareto efficient outcome in every subgame.

We modify the bargaining game described in Section 2 by transforming it into a game of *imperfect recall*.<sup>18</sup> At the end of each round of negotiation, we introduce a positive probability that the players might *forget* the previous history of play. It should be noticed that in the event of forgetfulness, we do allow the players to condition their future actions on time. In other words, the players forget the outcome path that has taken place so far, but are *not* constrained to play the same strategy starting at every forget information set.<sup>19</sup>

For reasons of tractability, the bargaining model with participation costs and imperfect recall that we analyze below is streamlined in various ways. As a first shot at modeling renegotiation within the extensive form of a game, we take forgetting as an entirely exogenous event rather than a strategic choice, and we assume that both players necessarily forget simultaneously the history of play.

<sup>18</sup> To our knowledge, bargaining games with imperfect recall have not been analyzed before in any form (see footnote 21 below for further references on games with imperfect recall). Chatterjee and Sabourian (2000) analyze a bargaining game (with  $N$  players) in which the players have *bounded memory* because of *complexity* considerations.

<sup>19</sup> Notice that imposing that the players play the same strategy at every possible forget information set would clash with the alternating offers nature of the bargaining protocol, which we want to preserve. The players need to know, at least, whether  $n$  is odd or even in order to know whose turn it is to be a proposer in the bargaining game.

We feel that while relaxing these assumptions would be desirable, this would be beyond the scope of this paper. We view Theorem 7 simply as a first step indicating that bounded recall may be a fruitful way to model renegotiation possibilities in an extensive form game.

Recall that, as we noted before, the crucial inefficient punishments in the bargaining game described in Section 2 are the ones used to punish a player who has not paid his participation cost. As with all off-the-equilibrium-path punishments these represent history dependent switches in the behavior of the players. The probability of forgetting the past history of play represents a chance to ‘forgive and forget’ for the players. More specifically, given that the players know the date even when they forget, they are able to infer something about the previous history of play, even when they find themselves at a forget information set—namely that an agreement has not been reached so far. Crucially, however, they are unable to distinguish between the possible different reasons for the failure to reach an agreement. There are, of course, three possible such reasons: failure to pay the participation costs, a deviation at the offer stage, and a deviation at the response stage of the previous bargaining rounds. When they forget, our players will be unable to punish (or reward) in different ways for these three types of behavior. Notice further that one of these three types of deviations naturally implies a reward in an alternating offers bargaining game. When the proposing player deviates to offer a share of the pie to the responder that is lower than what it should be, the responder must be rewarded in the future with a payoff that is larger than the offer he rejected. The necessary reward in this case builds into the extensive form a robust reason to avoid punishments for all three types of deviation when the players forget the past history of play.

Theorem 7 below states that when the probability of forgetting the past history of play in each period is above a minimum cut-off value (strictly below one), then the only equilibrium outcome of our modified bargaining game with imperfect recall is for neither player to ever pay his participation cost, and hence that no agreement is ever reached. In our view, this confirms that, when renegotiation possibilities are introduced, regardless of the values of participation costs, the unique equilibrium outcome of our model is that an agreement is never reached. In the presence of transaction costs and renegotiation embedded in the bargaining procedure, the Coase theorem may fail in a very strong way: no agreement is ever reached, and the entire surplus fails to materialize.

### 6.2. *Bargaining With Imperfect Recall*

The game that we analyze here is a modification of the game described in Section 2 above along the following lines. At the beginning of each period  $n \geq 1$ , we add an additional stage, stage  $O$ , in which Nature makes a chance move. Nature’s draws are described by a sequence of i.i.d. random variables  $\Delta = \{\mathcal{D}^1, \mathcal{D}^2, \dots, \mathcal{D}^n, \dots\}$ . The realization of each of the  $\mathcal{D}^n$  is denoted by  $d^n$  and

takes one of two possible values:  $d^n = \mathcal{F}$  (for ‘forget’) with probability  $q$ , and  $d^n = \mathcal{R}$  (for ‘recall’) with probability  $1 - q$ .<sup>20</sup>

The players do not observe the outcome of  $\mathcal{D}^n$  until the end of period  $n$ , after the responder has accepted or rejected the proposer’s offer in stage III of period  $n$  or either player has not paid his cost in stage I of period  $n$ . If the realization of  $\mathcal{D}^n$  is  $\mathcal{R}$ , the game moves to period  $n + 1$  (unless, of course, an offer has been made and accepted in period  $n$ , in which case the game terminates) with all the nodes corresponding to different outcome paths within period  $n$  belonging to distinct information sets. If, on the other hand, the realization of  $\mathcal{D}^n$  is  $\mathcal{F}$ , and the game has not terminated in period  $n$ , the players forget the previous history of play. In other words, in this case, for both players, all the nodes corresponding to stage I of period  $n + 1$ , via any possible history of play up to and including the whole period  $n$ , are in the same information set. The description of the extensive form within stages I, II, and III of each period is exactly the same as for the model described in Section 2 above.

We want to characterize the Nash equilibria of this game of imperfect recall that satisfy *sequential rationality*. As it is well known, in general, in games of imperfect recall this can pose a variety of technical problems and questions of interpretation. Luckily, in the case at hand matters are considerably simpler than in the general case.<sup>21</sup>

Given that we are dealing with a game of *incomplete information*, our equilibrium concept is (the weakest version of) *Perfect Bayesian Equilibrium* (PBE hereafter).<sup>22</sup>

**DEFINITION 2:** A *PBE for our bargaining game with imperfect recall and participation costs* is a pair of strategies and a set of beliefs such that, at every information set, the strategies are optimal given beliefs and beliefs are obtained from equilibrium strategies and observed actions using Bayes’ rule whenever possible.

We are now ready to state formally our last result.

**THEOREM 7:** *Consider the alternating offers bargaining game with participation costs and imperfect recall described above. For any given pair of costs  $(c_A, c_B)$  there exists a  $\bar{q} < 1$ , that is independent of the discount factors  $\delta_A$  and  $\delta_B$ , with the*

<sup>20</sup> While independence of these random variables plays a role in the proof of Theorem 7 below, it is easy to show that the actual probability  $q$  could be made to depend on time without affecting our results.

<sup>21</sup> Recently, Piccione and Rubinstein (1997) have sparked a debate on the interpretation of certain games of imperfect recall. We refer to their work and to the other papers in the special issue of *Games and Economic Behavior* (1997) for further details and references. Here we simply notice that the game we are analyzing does not exhibit *absent-mindedness* in the sense that they specify.

<sup>22</sup> See Fudenberg and Tirole (1991).

following property.<sup>23</sup> Whenever the probability  $q$  that the players forget the past history of play in every period exceeds  $\bar{q}$ , then the unique PBE outcome of the game is such that, along any possible realization of Nature's moves, both players never pay their participation cost in any period, and therefore an agreement is never reached.

The intuition behind the proof of Theorem 7 is relatively simple to describe. In a sense, it is a rather more complex version of the sunk-cost argument that provides the intuition for Theorem 6.

Suppose that an agreement  $x_B$  is reached in a period in which  $B$  is the proposer and  $A$  is the responder. The share  $x_B$  must satisfy several constraints. First of all,  $A$ 's net payoff,  $x_B - c_A$ , must be at least as much as what  $A$  gets if he does not pay his participation cost. This of course means that  $x_B$  must be at least as much as  $A$ 's continuation payoff if he does not pay his cost, plus  $c_A$ .

The agreed share  $x_B$  must also be less than or equal to the continuation payoff that  $A$  gets if he rejects offers below  $x_B$  in stage III of the agreement period. This is because  $A$  must be better off by rejecting *any* offer below  $x_B$  rather than by accepting it.

Putting the above two facts together tells us the following. The continuation payoff to  $A$  after he rejects offers below  $x_B$  cannot be smaller than  $A$ 's continuation payoff if he does not pay his cost in stage I, plus  $c_A$ . But, when the players forget the history, these two continuation payoffs for  $A$  must in fact be the *same*. Clearly this cannot be the case for large enough  $q$ , whenever  $c_A$  is positive. For large enough  $q$ , when  $c_A$  is positive,  $A$  is better off by not paying his participation cost, thus moving the game into the next period.

We view Theorem 7 as the most serious indication that inefficiencies are pervasive in our bargaining model with participation costs. In the original game that we described in Section 2 above, the no agreement equilibrium outcome, for low enough participation costs, was one of many possible ones. When the parties are given the possibility to renegotiate out of inefficient punishments, it is the only one that survives, for *any* positive values of the participation costs. In a bargaining game with positive participation costs, Coasian renegotiation opportunities destroy the efficient equilibria that a simple-minded interpretation of the Coase theorem would lead us to select among the many possible ones.

We conclude this Section with the observation that Theorem 7 directly implies the following. If we restrict attention to *stationary* strategies, the unique (Markov) Perfect Equilibrium outcome of the bargaining game with participation costs and perfect recall described in Section 2 above is that neither player ever pays his participation costs, and hence agreement is never reached. Indeed, Theorem 7 obviously guarantees that when  $q = 1$  the unique equilibrium outcome is that the costs are never paid and agreement is never reached. Notice that, setting

<sup>23</sup> In Remark A.1 we show that the bound  $\bar{q}$  can be made tighter if one is willing to make it dependent on the players' discount factors. We view the numerical value of these bounds as not particularly significant. In our view, what is important here is the qualitative nature of our inefficiency result.

$q = 1$  in the game with imperfect recall analyzed here is equivalent to restricting attention to strategies that can depend on the date, but not on the history of play in any previous bargaining rounds. This is a weaker restriction than imposing stationary strategies in the game with perfect recall. Hence the set of Perfect Equilibrium outcomes with stationary strategies in the game with perfect recall cannot be larger than what Theorem 7 predicts for the case of  $q = 1$ .

## 7. CONCLUSIONS

This paper shows that when negotiation takes place in the presence of transaction costs the Coase theorem does not necessarily hold. In particular we show that in an alternating offers bargaining game under perfect information, and with discounting, several types of inefficiencies may arise.

These inefficiencies should be viewed as pervasive for at least two reasons. First of all, we have shown that it is impossible consistently to select for efficient equilibria in our model. Secondly, and in our view more importantly, if the parties are given sufficient opportunities to break out of inefficient outcomes, the only outcome that survives in equilibrium is in fact the most inefficient possible one.

*Dept. of Economics, University of Southampton, Highfield, Southampton S017 1BJ, U.K.; and Dept. of Economics, Georgetown University, Washington, D.C. 20057, U.S.A.*

and

*Dept. of Economics, London School of Economics, Houghton St., London WC2A 2AE, U.K.*

*Manuscript received October, 1998; final revision received February, 2000.*

## APPENDIX

LEMMA A.1: *Consider the alternating offers bargaining game with participation costs described in Section 2. Whatever the values of  $\delta_i$  and  $c_i$  for  $i \in \{A, B\}$ , in any SPE of the game the payoffs to both players are nonnegative.*

Simply notice that either player can guarantee a payoff of zero by playing a strategy that prescribes never to pay any of his participation costs. *Q.E.D.*

PROOF OF THEOREM 1: We simply display a pair of strategies  $(\sigma_A^0, \sigma_B^0)$  that constitute an SPE of the game and that yield the desired outcome path.

For all  $i \in \{A, B\}$ , the strategy  $\sigma_i^0$  is described as follows. In stage I of any period  $\sigma_i^0$  prescribes that  $i$  does not pay his participation cost, regardless of the previous history of play. In stage II of any period in which  $i$  is a proposer,  $\sigma_i^0$  prescribes that  $i$  demands the entire pie for himself ( $x = 1$  if  $i = A$  and  $x = 0$  if  $i = B$ ), regardless of the previous history of play. In stage III of any period in which  $i$  is a responder,  $\sigma_i^0$  prescribes that  $i$  accepts any offer  $x \in [0, 1]$ , regardless of the previous history of play. It is easy to check that these strategies constitute an SPE of the game, and therefore this is enough to prove the claim. *Q.E.D.*

LEMMA A.2: Consider the alternating offers bargaining game with participation costs described in Section 2. Assume that  $\delta_i$  and  $c_i$  for  $i \in \{A, B\}$  are such that the game has an SPE in which an agreement is reached in finite time (see Theorem 2).

Let  $x_i^L$  be the infimum and  $x_i^H$  the supremum of all possible equilibrium agreements over the set of SPE in which an agreement is reached in a period in which player  $i$  is the proposer (the set of  $i$  SPE). Both  $x_i^L$  and  $x_i^H$  are undefined if the set of  $i$  SPE is empty.

Define also  $\tilde{x}_i^L$  and  $\tilde{x}_i^H$  to be the infimum and the supremum of all possible equilibrium agreements in a subgame in which, if the costs are paid,  $i$  is the first proposer (the  $i$  subgames). Both  $\tilde{x}_B^L$  and  $\tilde{x}_B^H$  are undefined if no equilibrium agreement is possible in a  $B$  subgame.<sup>24</sup>

Then  $\tilde{x}_i^L$  and  $\tilde{x}_i^H$ , and  $x_i^L$  and  $x_i^H$  are defined for all  $i \in \{A, B\}$ , and they satisfy  $x_A^H \leq 1 - c_B$ ,  $x_B^L \geq c_A$  as well as

$$(A.1) \quad x_B^H \leq \delta_A(x_A^H - c_A)$$

and

$$(A.2) \quad x_A^L \geq 1 - \delta_B(1 - x_B^L - c_B).$$

PROOF: By Lemma A.1, in any SPE the payoffs to both players must be nonnegative. The fact that it must be that  $\tilde{x}_A^H \leq 1 - c_B$  and  $\tilde{x}_B^L \geq c_A$  is now obvious since if the first inequality is violated  $B$  would get a negative payoff in some SPE and if the second inequality is violated,  $A$  would get a negative payoff in some (continuation) SPE.

By hypothesis, the set of SPE that prescribe some agreement is not empty. Therefore, either the set of  $A$  SPE is not empty, or the set of  $B$  SPE is not empty, or both are not empty.

If the set of  $B$  SPE is not empty we must have that

$$(A.3) \quad x_B^H \leq \delta_A(\tilde{x}_A^H - c_A).$$

To see this, consider the subgame that starts in stage III of the agreement period. If  $A$  rejects  $B$ 's offer at this stage, he will get a continuation payoff that is bounded above by  $\delta_A(\tilde{x}_A^H - c_A)$ . Therefore, it must be that  $A$ 's SPE strategy prescribes to accept *any* offer above  $\delta_A(\tilde{x}_A^H - c_A)$ . Therefore, in stage II of this period,  $B$ 's equilibrium strategy cannot be one that offers any  $x > \delta_A(\tilde{x}_A^H - c_A)$ , since otherwise he could reduce his offer by a small amount and  $A$  would still respond by accepting the offer. Therefore  $B$ 's offer must be some  $x \leq \delta_A(\tilde{x}_A^H - c_A)$ , and this is clearly enough to prove that (A.3) must hold in this case.

Notice next that (A.3) also implies the following. If the set of  $B$  SPE is not empty, then the set of  $A$  SPE is also not empty. This is because (A.3) says that  $x_B^H < \tilde{x}_A^H$  so that it must be the case that  $x_A^H = \tilde{x}_A^H$ . Therefore (A.3) proves (A.1).

Assume now that the set of  $A$  SPE is not empty. Our next step is to argue that  $\tilde{x}_B^L$  is defined in this case. Assume, by way of contradiction that it is not defined. Then an argument roughly symmetric to the one we used to show (A.3) proves that it would have to be the case that  $1 - x_A^L \leq 0$ . But this is a contradiction since it implies that  $x_A^L = 1$  and therefore that  $B$  should get a negative payoff in some SPE.

Using a completely symmetric argument to the one that proves that (A.3) must hold, we can now argue that if the set of  $A$  SPE is not empty, then we must have that

$$(A.4) \quad 1 - x_A^L \leq \delta_B(1 - \tilde{x}_B^L - c_B).$$

<sup>24</sup> It is worth it to clarify the differences between  $\tilde{x}_i^L$  and  $\tilde{x}_i^H$ , and  $x_i^L$  and  $x_i^H$  at this point. For instance,  $\tilde{x}_A^H$  is defined by assumption since we are postulating that there exists an SPE of the entire game with agreement in finite time. However, it *could* in principle be the case that  $x_A^H$  is undefined since all possible agreements could be reachable only in a period in which  $B$  is the proposer. As it turns out (see the statement of the Lemma),  $x_A^H$  is in fact defined.

Using a symmetric argument again, we can then see that (A.4) implies that if the set of  $A$  SPE is not empty then it must be the case that the set of  $B$  SPE is not empty either. Indeed, it must be the case that  $x_B^L = \tilde{x}_B^L$ . Therefore (A.4) proves (A.2).

Since we have just argued that the sets of  $A$  and  $B$  SPE are either both empty or both not empty, and by hypothesis at least one is in fact not empty, this is enough to prove the claim. *Q.E.D.*

PROOF OF THE ‘ONLY IF’ PART OF THEOREM 2: Using Lemma A.2, we know that if the set of SPE in which an agreement is reached is not empty we must have that

$$(A.5) \quad \begin{aligned} x_A^H &\leq 1 - c_B, & x_A^L &\geq 1 - \delta_B(1 - c_A - c_B), \\ x_B^H &\leq \delta_A(1 - c_A - c_B), & x_B^L &\geq c_A. \end{aligned}$$

Recalling that, by definition,  $x_i^H \geq x_i^L$  for  $i \in \{A, B\}$ , (A.5) directly implies (2). This is clearly enough to prove the claim. *Q.E.D.*

LEMMA A.3: Consider the alternating offers bargaining game with participation costs described in Section 2. Let  $\{x^n\}_{n=1}^\infty$  be a sequence of numbers such that  $x^n \in [c_A, 1 - c_B]$  for all  $n$  and such that for all odd  $n$

$$(A.6) \quad \delta_B(1 - x^{n+1} - c_B) \geq 1 - x^n$$

and for all even  $n$

$$(A.7) \quad \delta_A(x^{n+1} - c_A) \geq x^n.$$

Then there exists an SPE of the game  $(\sigma_A, \sigma_B)$  as follows.

(i) If at any point in the previous history of play either or both players have not paid their participation costs, then the strategies  $(\sigma_A, \sigma_B)$  revert to being the same as the strategies  $(\sigma_A^0, \sigma_B^0)$  described in the proof of Theorem 1 for the remainder of the game.

(ii) Unconditionally in stage I of period 1, and in stage I of every period conditionally on the fact that (i) above must not apply, both players pay their participation costs.

(iii) Provided that (i) above does not apply, in stage II of every period  $n$  the proposing player makes an offer  $x^n$  to the responding player.

(iv) Provided that (i) does not apply, in stage III of every period  $n$  the responding player accepts all offers that leave him with a share of the pie at least as large as the offer  $x^n$ , and he rejects all other offers.

(v) If the responding player rejects any offer that he is supposed to accept according to (iv) above, then the strategies  $(\sigma_A, \sigma_B)$  revert to being the same as the strategies  $(\sigma_A^0, \sigma_B^0)$  described in the proof of Theorem 1 for the remainder of the game.

PROOF: By Theorem 1, the strategies  $(\sigma_A, \sigma_B)$  constitute an equilibrium for any subgame following a history as in (i).

We now concentrate on the subgames starting stage I of an odd period  $n$ , following a history to which (i) does not apply (or the empty history if  $n = 1$ ). The argument for the even periods is symmetric and we omit the details.

Consider then any such subgame. By deviating and not paying his cost each player would earn a continuation payoff of zero. Following the prescription of  $(\sigma_A, \sigma_B)$  both players earn a continuation payoff of at least zero in any subgame. Therefore neither player has an incentive to deviate in any of these subgames.

Next, consider the subgame following the one above, starting in stage II of an odd period  $n$ . Clearly player  $A$  does not want to deviate and offer an  $x < x^n$  (the offer will be accepted and this will lower  $A$ 's payoff). Suppose now that player  $A$  deviates and offers  $x > x^n$ . Then his continuation payoff is  $\delta_A(x^{n+1} - c_A)$ . Since (A.6) implies  $x^n \geq x^{n+1}$ , we have that  $x^n > \delta(x^{n+1} - c_A)$ . Therefore, this is not a profitable deviation for player  $A$ .



Move now to the subgame following the one above, starting in stage III of an odd period  $n$ . At this point, some offer  $x$  has been made by  $A$ . Suppose first that  $x \leq x^n$ . At this point  $B$  is supposed to accept the offer  $x$ , and hence gets a continuation payoff of  $1 - x > 0$ . If  $B$  rejects the offer his continuation payoff is zero. Therefore this is not a profitable deviation for  $B$ . Suppose now that  $A$  has made an offer  $x > x^n$ , which  $B$  is supposed to reject. If  $B$  rejects, his continuation payoff is  $\delta_B(1 - x^{n+1} - c_B)$ . If  $B$  accepts, his continuation payoff is  $1 - x < 1 - x^n$ . But, using (A.6), we know that  $\delta_B(1 - x^{n+1} - c_B) \geq 1 - x^n$ . It follows that accepting the offer  $x$  is not a profitable deviation for  $B$ .

Therefore, no player has a profitable deviation from the behavior prescribed by  $(\sigma_A, \sigma_B)$  in any possible subgame. This is clearly enough to prove the claim. Q.E.D.

LEMMA A.4: Consider the alternating offers bargaining game with participation costs described in Section 2. Let  $\{x^n\}_{n=1}^\infty$  be a sequence of numbers in  $[c_A, 1 - c_B]$ , satisfying (A.6) and (A.7) as in Lemma A.3. Then, for every  $n$  odd, every  $A$  subgame has an SPE in which agreement is reached immediately and the agreed share of the pie is  $x^n$ , and for every  $n$  even, every  $B$  subgame has an SPE in which agreement is reached immediately and the agreed share of the pie is  $x^n$ .

PROOF: The claim is immediate using the strategies described in the proof of Lemma A.3. Q.E.D.

PROOF OF THE ‘IF’ PART OF THEOREM 2: It is enough to note that if  $\delta_i$  and  $c_i$  for  $i \in \{A, B\}$  are such that (2) holds, then (A.6) and (A.7) must hold when we set  $x^n = 1 - c_B$  for all odd  $n$  and  $x^n = c_A$  for all even  $n$ . Therefore the game has an SPE with immediate agreement as in Lemma A.4. This is enough to prove the claim. Q.E.D.

PROOF OF THE ‘IF’ PART OF THEOREM 3: Fix any  $x_A$  as in (5). Notice next that for such  $x_A$ , if we choose  $x^1 = x_A$ ,  $x^n = c_A$  for all even  $n$ , and  $x^n = 1 - c_B$  for all odd  $n \geq 3$ , we have a sequence  $\{x^n\}_{n=1}^\infty$  that satisfies (A.6) and (A.7) of Lemma A.3. By Lemma A.4, this is enough to prove the claim for the  $A$  subgames.

Symmetrically, now fix any  $x_B$  as in (6). Notice next that for such  $x_B$ , if we choose  $x^n = 1 - c_B$  for all  $n$ ,  $x^2 = x_B$  and  $x^n = c_A$  for all even  $n \geq 4$ , we have a sequence  $\{x^n\}_{n=1}^\infty$  that satisfies (A.6) and (A.7) of Lemma A.3. By Lemma A.4, this is enough to prove the claim for the  $B$  subgames. Q.E.D.

PROOF OF THE ‘ONLY IF’ PART OF THEOREM 3: If an SPE for an  $A$  subgame (a  $B$  subgame) were to exist, with immediate agreement on a share  $x_A$  (a share  $x_B$ ) outside the interval (5) (outside the interval (6)), we would have an immediate contradiction of (A.5). Q.E.D.

PROOF OF THEOREM 4: We concentrate on the claim for the payoffs of the  $A$  subgames. The argument for the  $B$  subgames is symmetric and therefore the details are omitted.

Let any  $x_A$  as in (5) and any  $n$  odd be given. We now display a pair of strategies  $(\sigma_A^A, \sigma_B^A)$  that constitute an SPE of the  $A$  subgames, and that give the players payoffs as in (7).

Up to and including period  $n - 1$  the strategies  $(\sigma_A^A, \sigma_B^A)$  are exactly the same as the strategies  $(\sigma_A^0, \sigma_B^0)$  of the proof of Theorem 1.

If any deviation from the prescribed outcome path is observed in any period  $1, \dots, n - 1$ , then the strategies  $(\sigma_A^A, \sigma_B^A)$  are again the same as the strategies  $(\sigma_A^0, \sigma_B^0)$  for the remainder of the game.

If no deviation from the prescribed outcome path is observed up to and including period  $n - 1$ , then the strategies  $(\sigma_A^A, \sigma_B^A)$  from stage I of period  $n$  are the same as the strategies  $(\sigma_A, \sigma_B)$  of the proof of the ‘if’ part of Theorem 3. Thus, the strategies in this subgame are SPE by construction, and yield an agreement of  $x_A$  in period  $n$ .

Next, let  $x_B$  as in (5) and any  $n$  even be given. As before, we display a pair of strategies  $(\sigma_A^A, \sigma_B^A)$  that constitute an SPE of the  $A$  subgames, and that give the players payoffs as in (8).

Up to and including period  $n - 1$  the strategies  $(\sigma_A^A, \sigma_B^A)$  are exactly the same as the strategies  $(\sigma_A^0, \sigma_B^0)$  of the proof of Theorem 1.

If any deviation from the prescribed outcome path is observed in any period  $1, \dots, n - 1$ , then the strategies  $(\sigma_A^A, \sigma_B^A)$  are again the same as the strategies  $(\sigma_A^0, \sigma_B^0)$  for the remainder of the game.

If no deviation from the prescribed outcome path is observed up to and including period  $n - 1$ , then the strategies  $(\sigma_A^A, \sigma_B^A)$  from stage  $I$  of period  $n$  are the same as the strategies  $(\sigma_A, \sigma_B)$  of the proof of the 'if' part of Theorem 3, starting in period 2. Thus, the strategies in this subgame are SPE by construction, and yield an agreement of  $x_B$  in period  $n$ . This is clearly enough to prove our claim. *Q.E.D.*

PROOF OF REMARK 1: Let  $(\sigma_A^N, \sigma_B^N)$  be an SPE of  $\Gamma^N$ . We concentrate on the case in which  $N$  is odd. The details for the case of  $N$  even are symmetric and hence they are omitted. We start by showing that  $(\sigma_A^N, \sigma_B^N)$  must prescribe that in stage I of period  $N$  neither player pays his participation cost, and therefore that the continuation payoffs to both players from the beginning of period  $N$  must be both 0.

Consider stage II of period  $N$ . By subgame perfection it is clear that  $A$  must make an offer  $x = 1$  to  $B$  at this stage. This is so because if  $B$  rejects  $A$ 's offer at this stage he earns a continuation payoff of zero, and hence his strategy must be to accept any  $x > 0$ . Therefore  $B$ 's continuation payoff in stage II of period  $N$  must be zero. It follows that if  $B$  pays  $c_B$  in stage I of period  $N$  his continuation payoff is  $-c_B$ . Clearly if he does not pay  $c_B$  at this stage he will earn a continuation payoff zero. Therefore,  $(\sigma_A^N, \sigma_B^N)$  must prescribe that  $B$  does not pay his participation cost in stage I of period  $N$ , and hence that  $A$  does not pay his cost either.

Once we know that the continuation payoffs for both players starting in stage I of period  $N$  are both zero we can move to stage I of period  $N - 1$ . Repeating the argument in the previous paragraph, with the players' roles exchanged, is now enough to show that  $(\sigma_A^N, \sigma_B^N)$  must prescribe that neither player pays his participation cost in stage I of period  $N - 1$ .

Continuing backwards up to stage I of period 1 is now enough to prove the claim. *Q.E.D.*

PROOF OF REMARK 2: Given any sequence  $\mathcal{S}$ , it is immediate to check that the strategies  $(\sigma_A^0, \sigma_B^0)$  of the proof of Theorem 1 constitute an SPE of  $\Gamma(\mathcal{S})$ . Since these strategies induce the required outcome path, this is enough to prove the claim. *Q.E.D.*

PROOF OF REMARK 3: The proof of Theorem 2 applies virtually unchanged. We do not repeat the details here. We only emphasize that Lemma A.3 and its proof apply to a bargaining game with participation costs and weakly alternating offers. In particular, the strategies described in the statement of Lemma A.3 do constitute valid strategies in this case. This is because, provided that clause (i) of Lemma A.3 does not apply, the only way for the game to proceed further into time is by a sequence of rejected offers. Therefore, since offers must alternate after a rejection, the identity of the proposer changes from each period to the next. *Q.E.D.*

LEMMA A.5: Consider the bargaining game with random proposer and participation costs described in Section 4. Assume that  $p$ ,  $\delta_i$  and  $c_i$ , for  $i \in \{A, B\}$ , are such that the game has an SPE in which an agreement is reached in finite time with positive probability (see the statement of Theorem 5).

Let  $x_{i,O}^L$  (respectively  $x_{i,E}^L$ ) be the infimum and  $x_{i,O}^H$  (respectively  $x_{i,E}^H$ ) the supremum of all possible equilibrium agreements over the set of SPE in which an agreement is reached with player  $i \in \{A, B\}$  being the proposer in an odd (respectively even) period. Let also  $x_O^L$ ,  $x_O^H$ ,  $x_E^L$ , and  $x_E^H$  be defined as follows:

$$(A.8) \quad \begin{aligned} x_O^L &= px_{A,O}^L + (1-p)x_{B,O}^L, & x_O^H &= px_{A,O}^H + (1-p)x_{B,O}^H, \\ x_E^L &= (1-p)x_{A,E}^L + px_{B,E}^L, & x_E^H &= (1-p)x_{A,E}^H + px_{B,E}^H. \end{aligned}$$

Both  $x_{i,k}^L$  and  $x_{i,k}^H$  for  $i \in \{A, B\}$  and  $k \in \{O, E\}$  are undefined if the corresponding set of SPE with agreement in finite time is empty.

If  $x_{i,k}^L$  and  $x_{i,k}^H$  are defined for all  $i \in \{A, B\}$ , and  $k \in \{O, E\}$ , then  $x_k^L \geq c_A$ ,  $x_k^H \leq (1 - c_B)$  for every  $k \in \{O, E\}$  as well as

$$(A.9) \quad x_{B,O}^H \leq \delta_A(x_E^H - c_A) \quad \text{and} \quad (1 - x_{A,O}^L) \leq \delta_B(1 - x_E^L - c_B)$$

and

$$(A.10) \quad x_{B,E}^H \leq \delta_A(x_O^H - c_A) \quad \text{and} \quad (1 - x_{A,E}^L) \leq \delta_B(1 - x_O^L - c_B).$$

If only  $x_{i,O}^L$  and  $x_{i,O}^H$  are defined for all  $i \in \{A, B\}$ , then they satisfy  $x_O^L \geq c_A$ ,  $x_O^H \leq (1 - c_B)$  as well as

$$(A.11) \quad x_{B,O}^H \leq \delta_A^2(x_O^H - c_A) \quad \text{and} \quad (1 - x_{A,O}^L) \leq \delta_B^2(1 - x_O^L - c_B).$$

Finally, if only  $x_{i,E}^L$  and  $x_{i,E}^H$  are defined for all  $i \in \{A, B\}$ , then  $x_E^L \geq c_A$ ,  $x_E^H \leq (1 - c_B)$  as well as

$$(A.12) \quad x_{B,E}^H \leq \delta_A^2(x_E^H - c_A) \quad \text{and} \quad (1 - x_{A,E}^L) \leq \delta_B^2(1 - x_E^L - c_B).$$

PROOF: We start by noticing that, in any SPE, each player's expected payoff at the beginning of each period must be nonnegative. Indeed, each player can guarantee himself an expected payoff of zero by following a strategy that prescribes never to pay any of his participation costs.

This observation implies that if  $x_{i,k}^L$  and  $x_{i,k}^H$  exist, then necessarily  $x_k^L \geq c_A$  and  $x_k^H \leq 1 - c_B$  for every  $k \in \{O, E\}$ . This must be the case since otherwise the expected equilibrium payoffs of one of the players would be negative in some SPE.

Assume now that  $x_{i,k}^L$  and  $x_{i,k}^H$  are defined for every  $i \in \{A, B\}$  and every  $k \in \{O, E\}$ . Consider an odd period; then necessarily

$$(A.13) \quad x_{B,O}^H \leq \delta_A(x_E^H - c_A).$$

Indeed, by rejecting  $B$ 's offer, the highest expected payoff that  $A$  can guarantee himself in the next (even) period is  $\delta_A(x_E^H - c_A)$ . This implies that any offer in excess of this payoff cannot be an equilibrium offer since  $A$  will certainly accept it and hence  $B$  can profitably deviate by reducing it while still guaranteeing acceptance. A symmetric argument proves the remaining inequality in (A.9).

A symmetric argument proves the two inequalities in (A.10). The details are omitted.

Assume now that only  $x_{i,O}^L$  and  $x_{i,O}^H$  are defined. Consider an odd period. Then an agreement cannot be reached in the following (even) period; therefore we can show that

$$(A.14) \quad x_{B,O}^H \leq \delta_A^2(x_O^H - c_A).$$

Indeed, since no agreement is reached in the following even period the highest expected payoff  $A$  can guarantee himself in the future is  $\delta_A^2(x_O^H - c_A)$  with agreement in the next odd period. Once again no offer in excess of this payoff can be an equilibrium offer since  $B$  can profitably reduce it, still guaranteeing acceptance. A symmetric argument can be used to prove the second inequality in (A.11).

Finally, in the case in which only  $x_{i,E}^L$  and  $x_{i,E}^H$  are defined, a symmetric argument proves the two inequalities in (A.12). Q.E.D.

LEMMA A.6: Consider the bargaining game with random proposer and participation costs described in Section 4. Assume that the game has an SPE in which agreement is reached in period  $n$ , conditionally on  $i \in \{A, B\}$  being the randomly chosen proposer. Then the game also has an SPE in which agreement is reached in period  $n$ , regardless of the identity of the randomly chosen proposer.

In other words, the quantities  $x_{i,k}^L$  and  $x_{i,k}^H$  for  $i \in \{A, B\}$  and  $k \in \{O, E\}$  in Lemma A.5 have the following property. For every given  $k \in \{O, E\}$ , either  $x_{A,k}^L$ ,  $x_{A,k}^H$ ,  $x_{B,k}^L$ , and  $x_{B,k}^H$  are all defined, or none is defined.

PROOF: We concentrate on the case in which an agreement is reached in an odd period  $n$ , conditionally on  $A$  being the proposer. The other three cases can be treated symmetrically, and the details are therefore omitted.

Consider then a pair of SPE strategies  $(\sigma_A, \sigma_B)$  yielding the following outcome path. In period  $n$  both players pay their costs. If the draw of the randomization device is  $\alpha$ ,  $A$  makes an offer  $x_A^n$  to  $B$ , and  $B$  subsequently accepts the offer. If instead the draw of the randomization device is  $\beta$ , then  $B$  makes an offer  $x_B^n$  to  $A$ , which is subsequently rejected by  $A$ .

Let  $\Pi_i$ , with  $i \in \{A, B\}$ , denote player  $i$ 's expected continuation payoff in this equilibrium, in the subgame that starts after  $A$  rejects  $B$ 's offer of  $x_B^n$ . Notice that  $\Pi_i \geq 0$  for all  $i \in \{A, B\}$ . Moreover, we must clearly have that

$$(A.15) \quad \Pi_A + \Pi_B \leq 1 - c_A - c_B.$$

Now construct two new strategies  $(\sigma'_A, \sigma'_B)$  by modifying  $(\sigma_A, \sigma_B)$  as follows. For both players, the prescriptions of  $(\sigma'_A, \sigma'_B)$  are the same as those of  $(\sigma_A, \sigma_B)$  in all subgames except for the one that starts in stage II of period  $n$ , after the outcome of the randomization device has turned out to be  $\beta$ . In this subgame the strategies  $(\sigma'_A, \sigma'_B)$  prescribe the following. In stage II of period  $n$ ,  $B$  makes an offer of  $x = \Pi_A$  to  $A$ . In stage III of period  $n$ ,  $A$  accepts all offers  $x' \geq \Pi_A$  and rejects all other offers. Therefore, the game terminates at this point. If at either stage II or stage III of period  $n$  either player deviates in a way that makes the game not terminate, then  $(\sigma'_A, \sigma'_B)$  prescribe actions that yield the players expected continuation payoffs of  $\Pi_A$  and  $\Pi_B$  respectively.

We now observe that the strategies  $(\sigma'_A, \sigma'_B)$  are enough to prove our claim. To see this, notice first that they yield the desired outcome path. Agreement takes place in period  $n$ , regardless of the identity of the randomly drawn proposer. Therefore, it only remains to show that  $(\sigma'_A, \sigma'_B)$  constitute an SPE of the game. To establish this fact, we need to check that neither player wants to deviate in any of the subgames in stages I, II, and III of period  $n$ .

Clearly, in stage I of period  $n$  both players want to pay their participation costs. This is because  $(\sigma_A, \sigma_B)$ , which constitute an SPE of the game, prescribe that both players pay their participation costs in stage I of period  $n$ . Clearly the continuation payoffs to both players from not paying the costs are unchanged with the new strategies. Moreover, since  $\Pi_i \geq 0$  for all  $i \in \{A, B\}$ , the continuation payoffs from paying the costs are no less with the new strategies than with the old strategies  $(\sigma_A, \sigma_B)$ .

Consider now a deviation from the part of player  $B$  in stage II of period  $n$ , to making an offer  $x' > \Pi_A$ . Clearly  $A$  will accept the offer, and therefore  $B$ 's payoff will be lower in this case. Therefore this is not a profitable deviation for  $B$ . Next, consider  $B$  deviating to an offer  $x' < \Pi_A$ . In this case  $A$  will reject the offer, and therefore  $B$ 's continuation payoff will become  $\Pi_B$ . However, by making the equilibrium offer  $x = \Pi_B$ ,  $B$  earns a continuation payoff of  $1 - \Pi_A$ . Using (A.15), the latter is clearly greater than  $\Pi_B$ . Therefore this is not a profitable deviation for  $B$ .

Move now to stage III of period  $n$ . Consider a deviation by  $A$  to accepting an offer  $x' < \Pi_A$ , which he is supposed to reject. By accepting the offer  $A$ 's expected continuation payoff is obviously  $x'$ , while if he rejects his continuation payoff is  $\Pi_A$ . Therefore this is not a profitable deviation for  $A$ . Finally, consider  $A$  deviating to rejecting an offer  $x' \geq \Pi_A$ . If he rejects, his expected continuation payoff is  $\Pi_A$ , while if he accepts it is obviously  $x' \geq \Pi_A$ . Therefore this is not a profitable deviation by  $A$ . This is clearly enough to conclude the proof of our claim. Q.E.D.

PROOF OF THE 'ONLY IF' PART OF THEOREM 5: We distinguish three cases:  $x_{i,k}^L$  and  $x_{i,k}^H$  exist and are defined for every  $i \in \{A, B\}$  and every  $k \in \{O, E\}$ ; only  $x_{i,O}^L$  and  $x_{i,O}^H$  are defined for every  $i \in \{A, B\}$ ; and finally only  $x_{i,E}^L$  and  $x_{i,E}^H$  are defined for every  $i \in \{A, B\}$ .

Notice that, by Lemma A.6, we know that these three cases are exhaustive of all possibilities.

Assume that  $x_{i,k}^L$  and  $x_{i,k}^H$  exist for every  $i \in \{A, B\}$  and every  $k \in \{O, E\}$ . Recall that (A.9) and (A.10) of Lemma A.5 tell us that in this case it must be that

$$(A.16) \quad \begin{aligned} x_{B,O}^H &\leq \delta_A(x_E^H - c_A), \\ x_{B,E}^H &\leq \delta_A(x_O^H - c_A). \end{aligned}$$

By definition (A.8) of  $x_E^H$  and  $x_O^H$  we also know that

$$(A.17) \quad \begin{aligned} x_E^H &= (1-p)x_{A,E}^H + px_{B,E}^H \leq (1-p) + px_{B,E}^H, \\ x_O^H &= px_{A,O}^H + (1-p)x_{B,O}^H \leq p + (1-p)x_{B,O}^H. \end{aligned}$$

Substituting (A.16) into (A.17) and using  $x_O^H \geq c_A$ ,  $x_E^H \geq c_A$  (from Lemma A.5) we obtain

$$(A.18) \quad x_E^H \leq (1-p) + p\delta_A(x_O^H - c_A),$$

$$(A.19) \quad x_O^H \leq p + (1-p)\delta_A(x_E^H - c_A).$$

Substituting further (A.19) into (A.18) and using  $x_E^H \geq c_A$  (from Lemma A.5) we can conclude that

$$(A.20) \quad p\delta_A(p - c_A) \geq c_A - (1-p).$$

Further substituting  $x_E^H \geq c_A$  and  $x_O^H \leq 1 - c_B$  (from Lemma A.5) into (A.18) we now get

$$(A.21) \quad p\delta_A(1 - c_B - c_A) \geq c_A - (1-p).$$

Combining (A.20) and (A.21) yields the second inequality in (9).

A completely symmetric argument and the remaining two inequalities in (A.9) and (A.10) prove the first inequality in (9). The details are omitted.

Assume now that only  $x_{i,O}^L$  and  $x_{i,O}^H$  are defined for every  $i \in \{A, B\}$ . From Lemma A.5 we get

$$(A.22) \quad x_{B,O}^H \leq \delta_A^2(x_O^H - c_A),$$

$$(A.23) \quad 1 - x_{A,O}^L \leq \delta_B^2(1 - x_O^L - c_B).$$

From the definition of  $x_O^H$  and  $x_O^L$  we have

$$(A.24) \quad x_O^H = px_{A,O}^H + (1-p)x_{B,O}^H \leq p + (1-p)x_{B,O}^H,$$

$$(A.25) \quad x_O^L = px_{A,O}^L + (1-p)x_{B,O}^L \geq px_{A,O}^L.$$

Substituting (A.24) and (A.25) into (A.22) and (A.23), respectively, we obtain

$$(A.26) \quad (p - c_A) \geq px_{B,O}^H \geq 0,$$

$$(A.27) \quad c_B \leq (1-p)x_{A,O}^L \leq (1-p).$$

Conditions (A.26) and (A.27) imply (10).

Finally, if only  $x_{i,E}^L$  and  $x_{i,E}^H$  are defined for every  $i \in \{A, B\}$  a symmetric argument proves (11). Q.E.D.

LEMMA A.7: *Consider the bargaining game with random proposer and participation costs described in Section 4. Whatever the values of  $p$ ,  $\delta_i$  and  $c_i$ , for  $i \in \{A, B\}$ , there exists an SPE of the game in which neither player pays his participation costs in any period, and hence agreement is never reached.*

PROOF: A pair of SPE strategies  $(\sigma_A^0, \sigma_B^0)$  yielding the prescribed outcome can be constructed adapting the strategies described in the proof of Theorem 1 to the new game. The details are omitted. Q.E.D.

LEMMA A.8: *Assume that Condition 1 of Theorem 5 is satisfied. Then there exists an SPE of the bargaining game with random proposer and participation costs such that agreement is reached in period 1 with probability one.*

PROOF: We proceed by construction. Consider the pair of strategies  $(\sigma'_A, \sigma'_B)$  defined as follows.

(i) If at any point in the previous history of play either or both players have not paid their participation costs, then the strategies  $(\sigma'_A, \sigma'_B)$  revert to being the same as the strategies  $(\sigma_A^0, \sigma_B^0)$  of Lemma A.7 for the remainder of the game.

(ii) Unconditionally in stage I of period 1, and in stage I of every period conditionally on the fact that (i) above must not apply, both players pay their participation costs.

(iii) Provided that (i) above does not apply, in stage II of every period  $n$  the proposing player  $i \in \{A, B\}$  makes an offer  $x_i^n$  to the responding player.

(iv) Provided that (i) above does not apply, in stage III of every period  $n$  the responding player accepts all offers that leave him with a share of the pie at least as large as the offer  $x_i^n$ , and he rejects all other offers.

(v) If the responding player rejects any offer that he is supposed to accept according to (iv) above, then strategies  $(\sigma'_A, \sigma'_B)$  revert to being the same as the strategies  $(\sigma_A^0, \sigma_B^0)$  of Lemma A.7 for the remainder of the game.

Let now  $x_i^n$  for  $i \in \{A, B\}$  be such that: if  $n$  is odd  $x_A^n = \min\{(1 - c_B)/p, 1\}$  and  $x_B^n = 0$ ; if instead  $n$  is even  $x_A^n = 1$  and  $x_B^n = \max\{c_A - (1 - p)/p, 0\}$ .

We can now show that, for these values of  $x_i^n$ ,  $i \in \{A, B\}$ , the strategies  $(\sigma'_A, \sigma'_B)$  are an SPE of the bargaining game with random proposer and participation costs.

By Lemma A.7, the strategies  $(\sigma'_A, \sigma'_B)$  constitute an equilibrium for any subgame following a history as in (i).

We now concentrate on the subgames starting in stage I of an odd period  $n$ , following a history to which (i) does not apply (or the empty history if  $n = 1$ ). The argument for the even periods is symmetric and we omit the details.

Consider any such subgame. By deviating and not paying his cost each player would earn a continuation payoff of zero. Following the prescription of  $(\sigma'_A, \sigma'_B)$  both players earn an expected continuation payoff of at least zero in any subgame. Therefore neither player has an incentive to deviate in any of these subgames.

Next, consider the subgame following the one above, starting in stage II of an odd period  $n$ . Clearly if, according to the randomization device, it is  $A$ 's turn to make an offer,  $A$  does not want to deviate and offer an  $x < \min\{(1 - c_B)/p, 1\}$  (the offer will be accepted and this will lower  $A$ 's payoff). Suppose now that player  $A$  deviates and offers  $x > \min\{(1 - c_B)/p, 1\}$ . Then his continuation payoff is zero. Since  $\min\{(1 - c_B)/p, 1\} \geq 0$  we conclude that  $A$  does not want to deviate. Symmetrically if, according to the randomization device, it is  $B$ 's turn to make an offer,  $B$  does not want to deviate and offer  $x \geq 0$  (the offer will be accepted and this will lower  $B$ 's payoff). Therefore, this is not a profitable deviation for player  $B$ .

Move now to the subgame following the one above, starting in stage III of an odd period  $n$ . Assume that following the outcome of the randomization device, some offer  $x$  has been made by  $A$ . Suppose first that  $x \leq \min\{(1 - c_B)/p, 1\}$ . At this point  $B$  is supposed to accept the offer  $x$ , and hence gets a continuation payoff  $(1 - x) \geq \max\{c_B - (1 - p)/p, 0\}$ . If  $B$  rejects the offer his expected continuation payoff is zero. Therefore this is not a profitable deviation for  $B$ . Suppose now that  $A$  has made an offer  $x > \min\{(1 - c_B)/p, 1\}$ , which  $B$  is supposed to reject. If  $B$  rejects, his expected continuation payoff is  $\delta_B(\min\{1 - c_A, p\} - c_A)$ . If  $B$  accepts, his continuation payoff is  $1 - x < \max\{c_B - (1 - p)/p, 0\}$ . But, using the first inequality in (9), we know that  $\delta_B(\min\{1 - c_A, p\} - c_B) \geq \max\{c_B - (1 - p)/p, 0\}$ . It follows that accepting the offer  $x$  is not a profitable deviation for  $B$ .

Assume now that following the outcome of the randomization device, some offer  $x \in [0, 1]$  has been made by  $B$ . The equilibrium strategies at this point prescribe that  $A$  accepts any offer  $x \geq 0$ . Therefore if  $A$  sticks to his equilibrium strategy, his payoff is  $x \geq 0$ . If he rejects his payoff is zero. Therefore  $A$  has no profitable deviation at this stage.

No player has a profitable deviation from the behavior prescribed by  $(\sigma'_A, \sigma'_B)$  in any possible subgame. This is clearly enough to prove that  $(\sigma'_A, \sigma'_B)$  is an SPE of the game. Q.E.D.

LEMMA A.9: *Assume that Condition 2 of Theorem 5 is satisfied. Then there exists an SPE of the bargaining game with random proposer and participation costs such that agreement is reached in period 1 with probability one.*

PROOF: Consider the strategies  $(\sigma''_A, \sigma''_B)$  defined as follows. In period 1 both players pay their participation costs. Still in period 1, following the outcome of the randomization device, the proposing player demands the entire surplus for himself:  $x_A^1 = 1$  and  $x_B^1 = 0$ . In stage III of period 1, the responding player accepts any offer  $x \in [0, 1]$ .

If any deviation from the equilibrium path occurs at any point in period 1, the strategies  $(\sigma_A'', \sigma_B'')$  revert to being the same as the strategies  $(\sigma_A^0, \sigma_B^0)$  of Lemma A.7 for the remainder of the game.

From the beginning of period 2 onward, regardless of the previous history of play, the strategies  $(\sigma_A'', \sigma_B'')$  are, again, the same as  $(\sigma_A^0, \sigma_B^0)$  of Lemma A.7.

Given that Condition 2 of Theorem 5 holds it is immediate to check that the strategies  $(\sigma_A'', \sigma_B'')$  constitute an SPE of the game. The details are omitted. Q.E.D.

LEMMA A.10: *Assume that Condition 3 of Theorem 5 is satisfied. Then there exists an SPE of the bargaining game with random proposer and participation costs such that agreement is reached in period 2 with probability one.*

PROOF: Consider the strategies  $(\sigma_A''', \sigma_B''')$  described as follows. In period 1 neither player pays his participation costs. In stage II and III of period 1 (off-the-equilibrium-path) the proposer demands the entire surplus for himself and the responder accepts any offer.

In period 2 both players pay their participation costs. Still in period 2, following the outcome of the randomization device, the proposing player demands the entire surplus for himself:  $x_A^2 = 1$  and  $x_B^2 = 0$ . In stage III of period 2, the responding player accepts any offer  $x \in [0, 1]$ .

If any deviation from the equilibrium path occurs at any point in period 1 or 2, the strategies  $(\sigma_A''', \sigma_B''')$  revert to being the same as the strategies  $(\sigma_A^0, \sigma_B^0)$  of Lemma A.7 for the remainder of the game.

From the beginning of period 3 onward, regardless of the previous history of play, the strategies  $(\sigma_A''', \sigma_B''')$  are, again, the same as  $(\sigma_A^0, \sigma_B^0)$  of Lemma A.7.

Given that Condition 3 of Theorem 5 holds it is immediate to check that the strategies  $(\sigma_A''', \sigma_B''')$  constitute an SPE of the game. The details are omitted. Q.E.D.

PROOF OF THE 'IF' PART OF THEOREM 5: The claim follows immediately from Lemma A.8, Lemma A.9, and Lemma A.10. Q.E.D.

PROOF OF REMARK 4: Since (12) implies that Condition 3 of Theorem 5 holds, the game has an SPE with agreement reached in period 2 with probability one, as shown in Lemma A.10.

Assume now that the game also has an SPE with agreement reached in period 1 with positive probability. Then, using Lemma A.6  $x_{i,k}^H$  and  $x_{i,k}^L$  must be defined for every  $i \in \{A, B\}$  and every  $k \in \{O, E\}$ . Therefore, as in the proof of the 'only if' part of Theorem 5, we must have that

$$(A.28) \quad p\delta_B(\min\{1 - c_B, p\} - c_B) \geq c_B - (1 - p)$$

which implies that

$$(A.29) \quad p\delta_B(p - c_B) > c_B - (1 - p)$$

and hence contradicts (12). Q.E.D.

PROOF OF THEOREM 6: Suppose, by way of contradiction, that the set of CPESPE is not empty. Notice that every CPESPE must yield an agreement in every subgame, whenever this is reached. Otherwise, since the players discount the future at a positive rate, the outcome could not possibly be Pareto efficient in every possible subgame.

Let  $x_i^H$  and  $x_i^L$  for  $i \in \{A, B\}$  be the supremum and the infimum respectively of the possible agreements in  $A$  periods and in  $B$  periods, taken over the set of all possible CPESPE.

The next few steps in the proof parallel closely the proof of the main result in Shaked and Sutton (1984).

Start with an  $A$  subgame. Since in stage III of each subgame  $B$  accepts any offer  $x$  below  $\delta_B(1 - x_B^L - c_B)$ , using subgame perfection we must have that

$$(A.30) \quad 1 - x_A^L \leq \delta_B(1 - x_B^L - c_B).$$

Moreover, since in stage III of any  $A$  subgame  $B$  rejects any  $x$  such that  $1 - x < \delta_B(1 - x_B^H - c_B)$ , we must have that

$$(A.31) \quad 1 - x_A^H \geq \delta_B(1 - x_B^H - c_B).$$

Using a symmetric argument for the  $B$  subgames we find that

$$(A.32) \quad x_B^H \leq \delta_A(x_A^H - c_A)$$

and

$$(A.33) \quad x_B^L \geq \delta_A(x_A^L - c_A).$$

Substituting (A.30) into (A.33) we now find that

$$(A.34) \quad x_B^L \geq \frac{\delta_A[1 - \delta_B(1 - c_B) - c_A]}{1 - \delta_A \delta_B}.$$

Substituting (A.31) into (A.32) we also obtain that

$$(A.35) \quad x_B^H \leq \frac{\delta_A[1 - \delta_B(1 - c_B) - c_A]}{1 - \delta_A \delta_B}$$

so that clearly we must have

$$(A.36) \quad x_B = x_B^H = x_B^L = \frac{\delta_A[1 - \delta_B(1 - c_B) - c_A]}{1 - \delta_A \delta_B}.$$

Symmetrically, substituting (A.32) into (A.31) and then (A.33) into (A.30) we also find out that

$$(A.37) \quad x_A = x_A^H = x_A^L = \frac{1 - \delta_B(1 - c_B + \delta_A c_A)}{1 - \delta_A \delta_B}.$$

Finally, notice that (A.36) and (A.37) together imply that

$$(A.38) \quad x_B = \delta_A(x_A - c_A)$$

and

$$(A.39) \quad 1 - x_A = \delta_B(1 - x_B - c_B).$$

Recall now that since an agreement must be reached in every subgame, it must be the case that both players pay their participation costs in stage I of every period. Consider now stage I of any  $A$  period. If player  $B$  pays his participation cost he gets a continuation payoff of

$$(A.40) \quad 1 - x_A - c_B$$

while if  $B$  deviates and does not pay his participation cost he gets a continuation payoff equal to

$$(A.41) \quad \delta_B(1 - x_B - c_B)$$

but, using (A.39), it is immediate that the quantity in (A.41) exceeds the quantity in (A.40). Therefore  $B$  finds it profitable to deviate and not pay his participation cost in stage I of every  $A$  subgame.

Symmetrically, we can verify that in stage I of every  $B$  subgame,  $A$  will find it profitable to deviate and not pay his participation cost. This is because (A.38) implies that

$$(A.42) \quad x_B - c_A < \delta_A(x_A - c_A).$$

Therefore, we have concluded that in every CPESPE, both players would have an incentive to deviate from their equilibrium behavior. This contradiction is clearly enough to prove the claim that the set of CPESPE is empty. *Q.E.D.*



PROOF OF THEOREM 7: Fix a pair of costs  $c_A$  and  $c_B$ . Next, suppose, by way of contradiction, that for every  $q \in (0, 1)$  there exist a PBE of the alternating offers bargaining game with participation costs and imperfect recall in which the parties reach an agreement for at least one of the realizations of the sequence of moves of Nature  $\sigma_N$ .

Let  $n(\sigma_N)$  be the period in which this agreement is reached. We start by considering the case in which this period is even. The details for  $n$  odd are in fact symmetric. For the remainder of the proof, we denote  $n(\sigma_N)$  simply by  $n$  for ease of notation. Moreover, all notation pertaining to the players information sets will be suppressed since the actual information set reached at the beginning of period  $n$  plays no role in our argument.

By our contradiction hypothesis, in period  $n$  both players pay their costs in stage I,  $B$  makes an offer  $x_B$  to  $A$  in stage II, and  $A$  accepts this offer in stage III.

Recall that the equilibrium beliefs of both players are consistent with equilibrium strategies and with Bayes' rule in every PBE of this game. Therefore, at  $n$ , the players' beliefs are entirely pinned down by the objective probability distribution over Nature's future moves.

For  $x_B$  to be an equilibrium offer it needs to be optimal for  $B$  to make such an offer. In other words, it must not be possible for  $B$  to make a lower offer  $x < x_B$  that  $A$  accepts in stage III of period  $n$ . This implies that for any offer  $x < x_B$ ,  $A$  must be at least as well off by rejecting  $x$  than by accepting it. This is the same as saying that the expected continuation payoff to  $A$ ,  $\Pi_A^E(x)$ , if he rejects the offer must be at least as high as  $x$ . Therefore

$$(A.43) \quad x < \Pi_A^E(x), \quad \forall x < x_B,$$

which trivially implies that

$$(A.44) \quad x_B \leq \Pi_A^E = \sup_{0 \leq x \leq x_B} \Pi_A^E(x).$$

The term  $\Pi_A^E$  can be bounded above, focusing on whether  $d^n$  is equal to  $\mathcal{F}$  or  $\mathcal{R}$ . With probability  $(1 - q)$  (corresponding to  $d^n = \mathcal{R}$ )  $A$ 's continuation payoff is at most  $\delta_A(1 - c_A - c_B)$ . This is because agreement can be reached at the earliest in period  $n + 1$ , and both  $A$  and  $B$  must pay their participation costs in period  $n + 1$  for this to be the case.

With probability  $q$ ,  $A$ 's continuation payoff after he rejects in stage III of period  $n$  is what he obtains after the players forget the history of play (corresponding to  $d^n = \mathcal{F}$ ). Let  $A$ 's continuation payoff in this case be denoted by  $\Pi_A^E(\mathcal{F})$ . Therefore we can now write

$$(A.45) \quad \Pi_A^E \leq q\Pi_A^E(\mathcal{F}) + (1 - q)\delta_A(1 - c_A - c_B)$$

and, using (A.44), we get

$$(A.46) \quad x_B \leq q\Pi_A^E(\mathcal{F}) + (1 - q)\delta_A(1 - c_A - c_B).$$

By our contradiction hypothesis that  $x_B$  is agreed at  $n$ , it must also be the case that it is optimal for  $A$  to pay the cost in stage I of period  $n$ . This implies that the equilibrium share  $x_B$  less the cost  $c_A$  needs to be higher than the expected continuation payoff to  $A$  if he does not pay his cost. Let this continuation payoff be denoted by  $\hat{\Pi}_A^E$ . We therefore have that

$$(A.47) \quad x_B - c_A \geq \hat{\Pi}_A^E.$$

With probability  $(1 - q)$  (corresponding to  $d^n = \mathcal{R}$ ) the continuation payoff to  $A$  after he does not pay his cost in stage I of period  $n$  is the payoff he gets if both players recall the history of the game at the end of period  $n$ . Clearly, this payoff must be at least zero.

With probability  $q$  the players forget the history of play (corresponding to  $d^n = \mathcal{F}$ ). In this case the outcome path starting in stage I of period  $n + 1$  is independent of what happens during period  $n$ . In other words, in this case the continuation payoff to  $A$  if he does not pay his participation cost must be precisely  $\Pi_A^E(\mathcal{F})$  as defined above. We can now conclude that

$$(A.48) \quad \hat{\Pi}_A^E \geq q\Pi_A^E(\mathcal{F})$$

and therefore, using (A.47), we now have that

$$(A.49) \quad x_B \geq q\Pi_A^E(\mathcal{F}) + c_A.$$

Putting together (A.46) and (A.49) yields

$$(A.50) \quad q\Pi_A^E(\mathcal{F}) + c_A \leq x_B \leq q\Pi_A^E(\mathcal{F}) + (1-q)\delta_A(1-c_A-c_B),$$

which trivially implies that it must be the case that

$$(A.51) \quad c_A \leq (1-q)\delta_A(1-c_A-c_B).$$

Notice now that (A.51) is a contradiction unless

$$(A.52) \quad q \leq \frac{\delta_A(1-c_A-c_B) - c_A}{\delta_A(1-c_A-c_B)}.$$

Using a completely symmetric argument (the details are therefore omitted), it is possible to show that an agreement in any odd period  $n$  yields a contradiction unless

$$(A.53) \quad q \leq \frac{\delta_B(1-c_A-c_B) - c_B}{\delta_B(1-c_A-c_B)}.$$

Let now

$$(A.54) \quad \hat{q} = \max\left\{\frac{\delta_A(1-c_A-c_B) - c_A}{\delta_A(1-c_A-c_B)}, \frac{\delta_B(1-c_A-c_B) - c_B}{\delta_B(1-c_A-c_B)}\right\}$$

and

$$(A.55) \quad \bar{q} = \max\left\{\frac{1-2c_A-c_B}{1-c_A-c_B}, \frac{1-c_A-2c_B}{1-c_A-c_B}\right\}.$$

Notice that since  $\delta_A < 1$ ,  $\delta_B < 1$ ,  $c_A \in (0, 1)$ , and  $c_B \in (0, 1)$  we have that  $\hat{q} < \bar{q} < 1$ . Since any agreement for any  $q > \hat{q}$  yields a contradiction, this is clearly enough to prove the claim. *Q.E.D.*

REMARK A.1: *The bound  $\bar{q}$  used in Theorem 7 as in (A.55) can, in general be made tighter at the cost of making it dependent of the players' discount factors. A tighter lower bound on  $q$  that ensures that no equilibrium with agreement exists is given by  $\hat{q}$  as in (A.54) above.*

## REFERENCES

- ABREU, D., D. PEARCE, AND E. STACCHETTI (1993): "Renegotiation and Symmetry in Repeated Games," *Journal of Economic Theory*, 60, 217–240.
- AGHION, P., AND J. TIROLE (1997): "Formal and Real Authority in Organizations," *Journal of Political Economy*, 105, 1–29.
- ANDERLINI, L., AND L. FELLI (1997): "Costly Coasian Contracts," C.A.R.E.S.S. Working Paper 97-11, University of Pennsylvania.
- BENOÎT, J.-P., AND V. KRISHNA (1993): "Renegotiation in Finitely Repeated Games," *Econometrica*, 61, 303–323.
- BERNHEIM, B. D., AND D. RAY (1989): "Collective Dynamic Consistency in Repeated Games," *Games and Economic Behavior*, 1, 295–326.
- BUSCH, L.-A., AND Q. WEN (1995): "Perfect Equilibria in a Negotiation Model," *Econometrica*, 63, 545–565.
- CHATTERJEE, K., AND H. SABOURIAN (2000): "Multiperson Bargaining and Strategic Complexity," *Econometrica*, 68, 1491–1509.

- COASE, R. H. (1960): "The Problem of Social Cost," *The Journal of Law and Economics*, 3, 1–44.
- (1992): "The Institutional Structure of Production," *American Economic Review*, 82, 713–719.
- DE MEZA, D. (1988): "Coase Theorem," in *The New Palgrave Dictionary of Economics and the Law*, ed. by P. Newmann. New York: Macmillan, pp. 270–282.
- DIXIT, A., AND M. OLSON (1997): "Does Voluntary Participation Undermine the Coase Theorem?" *Journal of Public Economics*, 76, 309–335.
- FARRELL, J., AND E. MASKIN (1987): "Renegotiation in Repeated Games," Discussion Paper 1335, Harvard Institute of Economic Research, Harvard University.
- (1989): "Renegotiation in Repeated Games," *Games and Economic Behavior*, 1, 327–360.
- FERNANDEZ, R., AND J. GLAZER (1991): "Striking for a Bargain Between Two Completely Informed Agents," *American Economic Review*, 81, 240–252.
- FERSHTMAN, C., AND D. SEIDMANN (1993): "Deadline Effects and Inefficient Delays in Bargaining with Endogenous Commitments," *Journal of Economic Theory*, 60, 306–321.
- FUDENBERG, D., AND J. TIROLE (1991): "Perfect Bayesian Equilibrium and Sequential Equilibrium," *Journal of Economic Theory*, 53, 236–260.
- GAMES AND ECONOMIC BEHAVIOR (1997): *Special Issue on Imperfect Recall*, Vol. 20, No. 1. New York: Academic Press.
- GROSSMAN, S. J., AND O. D. HART (1986): "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94, 691–719.
- GROUT, P. (1984): "Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Solution," *Econometrica*, 52, 449–460.
- HART, O. D., AND J. MOORE (1990): "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98, 1119–1158.
- KLEIN, B., R. G. CRAWFORD, AND A. A. ALCHIAN (1978): "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *Journal of Law and Economics*, 21, 297–326.
- MUTHOO, A. (1999): *Bargaining Theory with Applications*. Cambridge: Cambridge University Press.
- NICHOLSON, W. (1989): *Microeconomic Theory: Basic Principles and Extensions*, Fourth International Edition. Chicago: The Dryden Press.
- PICCIONE, M., AND A. RUBINSTEIN (1997): "On the Interpretation of Decision Problems with Imperfect Recall," *Games and Economic Behavior*, 20, 3–24.
- RAJAN, R., AND L. ZINGALES (1998): "Power in a Theory of the Firm," *Quarterly Journal of Economics*, 113, 387–432.
- RIEDL, A. (1997): "Bargaining and Opting In Costs," Institute for Advanced Studies, Department of Economics, Vienna, mimeo.
- RUBINSTEIN, A. (1982): "Perfect Equilibrium in a Bargaining Model," *Econometrica*, 50, 97–110.
- SHAKED, A., AND J. SUTTON (1984): "Involuntary Unemployment as a Perfect Equilibrium in a Bargaining Game," *Econometrica*, 52, 1351–1364.
- STÅHL, I. (1972): *Bargaining Theory*. Stockholm: Stockholm School of Economics.
- WILLIAMSON, O. (1985): *The Economic Institutions of Capitalism*. New York: Free Press.