

Coase Theorem, Complexity and Transaction Costs

Jihong Lee

Birkbeck College, London*

Hamid Sabourian

Birkbeck College, London and King's College, Cambridge[†]

March 2006

Abstract

This paper, by introducing complexity considerations, provides a dynamic foundation for the Coase Theorem and highlights the role of transaction costs in generating inefficient outcomes in bargaining/negotiation. We show that, when the players have a preference for less complex strategies, the Coase Theorem is valid in negotiation models with repeated surplus and endogenous disagreement payoffs if and only if there are no transaction costs. Specifically, complexity considerations allow us to select only *efficient* equilibria in these models without transaction costs while in sharp contrast every equilibrium outcome induces perpetual disagreement and *inefficiency* with (arbitrarily small) transaction costs. We also show that the latter is true in the Rubinstein bargaining model with transaction costs.

JEL Classification: C72, C78

Keywords: Coase Theorem, Efficiency, Bargaining, Repeated Game, Transaction Cost, Complexity, Bounded Rationality, Automaton

*School of Economics, Mathematics and Statistics, Birkbeck College, Malet St, London, WC1E 7HX, United Kingdom (email: J.Lee@econ.bbk.ac.uk)

[†]Faculty of Economics and Politics, Sidgwick Avenue, Cambridge, CB3 9DD, United Kingdom (email: Hamid.Sabourian@econ.cam.ac.uk)

1 Introduction

This paper, by introducing complexity considerations, explores the extent of the validity of the Coase Theorem. In particular, we highlight the role of “transaction costs” in explaining why individuals may not fully exploit mutual gains from trade via bargaining and negotiation. The central message of the paper is that, when each player has a preference for less complex strategies (at the margin), only *efficient* equilibria arise in complete information models of bargaining/negotiation without transaction costs while, in sharp contrast, perpetual disagreement, and *inefficiency*, are the only possible features of an equilibrium outcome with arbitrarily small transaction costs. Thus, in what follows the Coase Theorem is valid if and only if there are no transaction costs.

We consider two departures from the Rubinstein [17] two-player alternating-offers bargaining game that have been proposed to explain the possibility of Coasian failure under complete information. First, the “negotiation game” of Busch and Wen [5] (henceforth BW) modifies Rubinstein’s model in the following two ways: bargaining is over the distribution of a fixed *periodic* surplus and disagreement payoffs in each period are determined *endogenously* in some game. A well-known special case of this game is the standoff between a union and a firm considered by Fernandez and Glazer [9] and Haller and Holden [11]. Second, in their “costly bargaining game”, Anderlini and Felli [2] (henceforth AF) introduce transaction costs to the Rubinstein model; bargaining in each period takes place if and only if both players sink a (small) participation cost.

In contrast to the Rubinstein model, in BW, the forces of endogenous disagreement payoffs and periodic surplus, and in AF, the presence of transaction costs, induce a large number of equilibria, a result reminiscent of the Folk theorem of repeated games. Some of these equilibria involve delay in agreement (even perpetual disagreement) and inefficiency while some are efficient. Thus, the Coase Theorem fails in the sense that it is no longer guaranteed.

Such multiplicity of equilibria, however, makes it difficult to draw firm conclusions about the Coase Theorem. Moreover, one may ask which of the two features, endogenous disagreement payoffs and periodic surplus or transaction costs, plays a more critical role in driving inefficient allocations. Motivated by the recent literature on complexity and bargaining games which has yielded some sharp selection results (Chatterjee and Sabourian [6][7], Sabourian [19] and Gale and Sabourian [10]), we address these issues by introducing complexity costs of implementing strategies into the above games.

Our main analysis examines BW’s negotiation game without and with transaction costs. In the latter case, henceforth called the “costly negotiation game”, we marry AF with BW by assuming that in order for the players to bargain in each period of the negotiation game (but not to play the disagreement game) both have to pay a small participation cost; if at least one player foregoes the payment, there is no bargaining and they proceed directly to the disagreement game. We also briefly consider the Rubinstein

model with participation costs (the AF setup) and confirm our message.

The negotiation model considered in our main analysis can be interpreted either as a bargaining model as described above, or alternatively, as a repeated game with exit options via bargaining and contractual agreement. Repeated interactions are often accompanied by negotiations which can lead to mutual enforceable agreement. While equilibria in standard repeated games are usually given the interpretation of (implicit) self-enforcing contracts, the negotiation game depicts situations with the possibility of explicit contracts. For example, we observe firms engaged in a repeated horizontal or vertical relationship negotiating over a long-term contract, or even a merger; similarly, countries involved in international trade often attempt to settle an agreement that enforces fixed quotas and tariffs. From this perspective, our analysis can also be thought of as an investigation of the Coase Theorem and the role of transaction costs in the context of repeated games with exit options.

The key details of our complexity framework and results are as follows. We follow Kalai and Stanford [12] and define complexity of a strategy by the number of *continuation strategies* that the strategy induces at different periods/histories of the game.¹ The complexity analysis is then facilitated by considering equivalent “machine games”. Using a new machine specification that formally distinguishes between the two *roles* - proposer and responder - played by each player in a given period, we establish for the games considered the equivalence between Kalai and Stanford’s notion of complexity and the counting-the-number-of-states measure of complexity introduced in the literature on repeated games played by finite automata (see [18], [1], [15] and [16]). The concept of Nash equilibrium is refined by introducing complexity cost *lexicographically* with the standard payoff into the players’ preference orderings. We refer to it as a “Nash equilibrium of the machine game (NEM)”. We define a “subgame-perfect equilibrium of the machine game (SPEM)” as a NEM that is subgame-perfect.

Without transaction costs, complexity considerations select only efficient equilibria in the negotiation game under sufficiently patient players. More specifically, if an agreement occurs in some finite period as a NEM outcome then it must occur within the first two periods of the game. Thus, in this case the outcome is efficient in the limit as the discount factor goes to one. We then show that, given sufficiently patient players, every SPEM in the negotiation game that induces perpetual disagreement is at least *long-run* almost efficient; that is, the game must reach a finite date at which the continuation game then on is almost efficient. It also follows that, given sufficiently patient players, if every disagreement game outcome is inefficient then perpetual disagreement cannot occur, and thus, every SPEM induces an agreement in the first two periods and is almost efficient. Other restrictions on the disagreement game (for example, a unique Nash equilibrium,

¹Our complexity definition differs from that used in the aforementioned literature on complexity and bargaining ([6], [7], [19] and [10]) which focuses on complexity of the response rule within a period.

as in the union-firm negotiation models of [9] and [11]) ensure that every SPEM inducing an agreement in the first two periods is efficient independently of the discount factor.

Introducing transaction costs sharply alters the equilibrium picture from efficiency to inefficiency. In particular, we show that, for any discount factor and any transaction cost, every SPEM in the costly negotiation game induces either perpetual disagreement or an agreement within the first two periods. Furthermore, the former is the only feasible SPEM outcome if the disagreement game has a unique Nash equilibrium.² Thus, in such a case, any SPEM is inefficient if in addition every disagreement game outcome is Pareto-dominated by an agreement.

Finally, we discuss how our central message resonates beyond the chosen setup: complexity/machine specifications, equilibrium concepts and extensive form. In fact, by considering alternative machine specifications that account for finer partitions of histories and/or equilibrium concepts that involve a positive tradeoff between complexity and payoffs in preferences, the selection results, both without and with transaction costs, become sharper (for example, they no longer depend on the values of the discount factor and the transaction cost). We also confirm the dramatic impact of transaction costs in the Rubinstein bargaining model by applying complexity considerations to AF's "costly bargaining game". Here also, for any positive transaction cost and discount factor, it is only the most *inefficient* equilibrium involving perpetual disagreement that survives with complexity-averse players.

Our complexity analysis therefore suggests the following. First, transaction costs are a critical ingredient of a robust account of why perpetual disagreement and inefficient outcomes can arise in bargaining/negotiation. Second, in absence of transaction costs, the Coase Theorem can be extended to negotiation models in which the surplus is periodic and disagreement payoffs are endogenous. Also, since these models can alternatively be thought of as repeated games with exit options, this result takes the study of complexity in repeated games a step further from the existing literature in which complexity or bargaining alone has produced only a limited selection result.³ While many inefficient equilibria survive complexity refinement, we demonstrate that complexity and bargaining in tandem ensure efficiency in repeated interactions.

The paper is organized as follows. In Section 2, we first describe BW's negotiation game (without transaction costs); then introduce the notion of complexity in terms of strategies/machines, describe the machine game and present the efficiency results. In Section 3, we describe the costly negotiation game and apply the tools developed in the

²More generally, perpetual disagreement is the only feasible outcome if the transaction cost exceeds some parameter that is determined by the set of Nash equilibria of the disagreement game.

³Although complexity refinement narrows the set of equilibrium payoffs in 2×2 repeated games (see [1]), this is not the case more generally. For example, Bloise [4] shows robust examples of 2-player repeated games with pure strategies and patient players in which the set of Nash equilibrium payoffs with complexity costs coincides with the set of strictly individually rational payoffs.

previous section to demonstrate the dramatic impact of (arbitrarily small) transaction costs. In Section 4, we first discuss several channels via which the analysis can be extended and the results further sharpened. We then show how complexity also selects perpetual disagreement as the only equilibrium outcome in the costly bargaining game of AF. Except for those omitted to save space, all proofs are relegated to Appendix.

2 Negotiation Game and Complexity

2.1 The Model

Let us formally describe the negotiation game, as defined by BW. There are two players indexed by $i = 1, 2$. In the alternating-offers protocol, each player in turn proposes a partition of a *periodic* surplus whose value is normalized to one. If the offer is accepted, the game ends and the players share the surplus accordingly at every period indefinitely thereafter. If the offer is rejected, the players engage in a one-shot (normal form) game, called the “disagreement game”, before moving onto the next period in which the rejecting player makes a counter-offer.

We index the (potentially infinite) time periods by $t = 1, 2, \dots$ and adopt the convention that player 1 makes offers in odd periods and player 2 makes offers in even periods. Let $\Delta^2 \equiv \{x = (x_1, x_2) \mid \sum_i x_i = 1\}$ be a partition of the unit periodic surplus. A period then refers to a single offer $x \in \Delta^2$ by one player, a response made by the other player - acceptance “Y” or rejection “N” - and the play of the disagreement game if the response is rejection. The common discount factor is $\delta \in (0, 1)$.

The disagreement game is a normal form game, defined as $G = \{A_1, A_2, u_1(\cdot), u_2(\cdot)\}$, where A_i is the set of player i 's strategies (or simply actions) and $u_i(\cdot) : A_1 \times A_2 \rightarrow R$ is his payoff function. We shall denote the set of action profiles in G by $A = A_1 \times A_2$ with its element indexed by a . Let $u(\cdot) = (u_1(\cdot), u_2(\cdot))$ be the vector of payoff functions, and assume that it is bounded. Each player's minmax payoff in G is normalized to zero. Also, we assume that, for any $a \in A$, $u_1(a) + u_2(a) \leq 1$. Therefore, bargaining offers the players an opportunity to settle on an efficient outcome once and for all.

Next, let T denote the end of the negotiation game and $a^t \in A$ the disagreement game outcome (action profile) in period t . If the players reach an agreement on the partition $z = (z_1, z_2) \in \Delta^2$ at $T < \infty$, player i 's (discounted) *average* payoff in the negotiation game is $(1 - \delta) \sum_{t=1}^{T-1} \delta^{t-1} u_i(a^t) + \delta^{T-1} z_i$. If an agreement is never reached (we describe this by setting $T = \infty$) player i 's corresponding payoff is $(1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} u_i(a^t)$.

The negotiation game is stationary every two periods; henceforth, we refer to every two periods (beginning with an odd one) by a stage. In specifying strategies (and later machines), we formally distinguish between the two different *roles* - proposer (p) or responder (r) - played by each player in every stage. We index a player's role by k .

In order to define a strategy, we need further notation. First, whenever superscripts/subscripts i and j both appear in the same exposition, we mean $i, j = 1, 2$ and $i \neq j$. Similarly, whenever we use superscripts/subscripts k and l together, we mean $k, l = p, r$ and $k \neq l$.

Denote a history of outcomes in a period by e and the set of all such outcomes by

$$E = \{(x^i, Y), (x^i, N, a)\}_{x^i \in \Delta^2, a \in A, i=1,2},$$

where the partition x^i refers to an offer by player i . Let e^t be the outcome of period t .

We also need to represent information available to a player *within* a period when it is his turn to take an action given his role. To this end, let d be a “partial history” (information within a period) and define D_{ik} as the set of partial histories at which it is player i 's turn to play in role k . We then have

$$D_{ip} = \{\emptyset, (x^i, N)\}_{x^i \in \Delta^2} \text{ and } D_{ir} = \{(x^j), (x^j, N)\}_{x^j \in \Delta^2},$$

where the null set \emptyset refers to the beginning of a period at which the proposer has to make an offer, (x^i) is the partial history of offer x^i by player i and (x^i, N) is the partial history of offer x^i by i followed by the other player's rejection.

We denote the set of actions available to player i by $C_i \equiv \Delta^2 \cup Y \cup N \cup A_i$. Let $C_{ik}(d)$ be the set of actions available to i given his role k and a partial history $d \in D_{ik}$. Thus,

$$C_{ip}(d) = \begin{cases} \Delta^2 & \text{if } d = \emptyset \\ A_i & \text{if } d = (x^i, N) \end{cases} \text{ and } C_{ir}(d) = \begin{cases} \{Y, N\} & \text{if } d = x^j \\ A_i & \text{if } d = (x^j, N) \end{cases}.$$

Define H^t to be the set of all possible past histories of outcomes at date t in the negotiation game, excluding those that have resulted in an agreement. Thus, $H^t \subseteq E^{t-1}$ ($t-1$ -fold Cartesian product of E) for any $t > 1$ and $H^1 = \emptyset$ is the initial empty (trivial) history. Let $H^\infty \equiv \cup_{t=1}^\infty H^t$ denote the set of all possible finite period histories.

For the analysis, we divide H^∞ into two smaller subsets according to the different roles taken by the players in each stage. Let $H_{ik}^\infty \subset H^\infty$ be the set of all histories at which player i is in role k . Note that $H_{ik}^\infty = H_{jl}^\infty$ and $H^\infty = H_{ip}^\infty \cup H_{ir}^\infty$ for any $i = 1, 2$.

We assume that the players have perfect information and do not randomize. A strategy for player i is then a function $f_i : (H_{ip}^\infty \times D_{ip}) \cup (H_{ir}^\infty \times D_{ir}) \rightarrow C_i$ such that $f_i(h, d) \in C_{ik}(d)$ for any $(h, d) \in H_{ik}^\infty \times D_{ik}$. The set of all strategies for player i is denoted by F_i . Denote by F_i^t the set of player i 's strategies in the negotiation game starting with role distribution given in period t . Thus, if t is odd, $F_i^t = F_i$. Also, let $\pi_i(f)$ denote i 's payoff in the negotiation game under strategy profile $f = (f_1, f_2)$.

Next, we define a *stationary* (history-independent) strategy in the following way.

Definition 1 *A strategy f_i is stationary if and only if $f_i(h, d) = f_i(h', d)$ for any $h, h' \in H_{ik}^\infty$, any $d \in D_{ik}$ and any $k = p, r$. A strategy profile $f = (f_i, f_{-i})$ is stationary if and only if f_i is stationary for all i .*

Note that the behavior induced by such a strategy may depend on the partial history within the current period but not on the history of the game up to it. For instance, if a strategy f_i is stationary, then it must be such that $f_i(h, x, N) = f_i(h', x, N)$ for any h, h' and x ; but we may have $f_i(h, x, N) \neq f_i(h, x', N)$ for some x and x' .⁴ Note also that a stationary strategy profile always induces the same outcome in each stage of the game.

BW provide a characterization for the set of subgame-perfect equilibrium (SPE) payoffs of the negotiation game. They show that the forces of bargaining somewhat restrict the set of feasible equilibrium payoffs in the negotiation game *vis-à-vis* the set of individually rational payoffs in the disagreement game. However, as in the Folk Theorem of repeated games, the negotiation game, in general, still admits a large number of SPEs (even when all the disagreement payoffs are uniformly small relative to agreement).⁵

2.2 Complexity, Machines and Equilibrium

There are many alternative ways to think of the “complexity” of a strategy in dynamic games. One natural and intuitive way to measure strategic complexity, which we shall adopt in the paper, is to consider the total number of distinct *continuation strategies* that the strategy induces at different histories (Kalai and Stanford [12], henceforth KS).

In a repeated game in which the stage game is one-shot, it is natural to take the measure over all histories at the beginning of each period (and this coincides with the set of all subgames). In the negotiation game, as well as in the other games we consider below, every stage game consists of an extensive form game, and thus several different definitions are possible depending on the definitions of the set of histories. Here, we shall consider the set of all continuation strategies *at the beginning of each period*.

Formally, let $f_i|h$ be the continuation strategy at history $h \in H^\infty$ induced by $f_i \in F_i$. Thus, $f_i|h(h', d) = f_i(h, h', d)$ for any $h, h' \in H_{ik}^\infty$, any $d \in D_{ik}$ and any k . Define the set of all such continuation strategies by $F_i(f_i) = \{f_i|h : h \in H^\infty\}$. The cardinality of this set, referred to by $comp(f_i)$, shall provide a measure of strategic complexity.⁶

The set of continuation strategies can also be divided into smaller sets according to the role specification. Let $F_{ik}(f_i) = \{f_i|h : h \in H_{ik}^\infty\}$ be the set of all such continuation strategies for i in role k . Then $F_i(f_i) = \cup_k F_{ik}(f_i)$ and $comp(f_i) = \sum_k |F_{ik}(f_i)|$.

⁴BW define a stationary strategy, but their definition differs from ours in that it does not allow for disagreement game actions to depend on the partial history within a period (and thus has more of a Markovian flavor).

⁵BW show that there exists a profile of payoffs \underline{v} such that any payoff profile exceeding \underline{v} can be sustained by a SPE if players are sufficiently patient. They further show that, in the limit as δ tends to 1, the negotiation game has a unique efficient SPE payoff vector if and only if \underline{v} is efficient, which is true if and only if $\max_{a \in A} \{u_i(a) - [\max_{a'_j \in A_j} u(a'_j, a_{-j}) - u_j(a)]\} = 0$ for any $i \neq j$.

⁶We can also measure complexity over finer partitions of histories/subgames (including subgames starting within a period) and corresponding continuation strategies. We shall later discuss how we can sharpen our results by using such finer measures.

In the context of repeated games KS show that any strategy can always be represented by an automaton or a “machine” and that the above notion of complexity of a strategy (the number of continuation strategies) is equivalent to counting the number of states of the smallest automaton that implements the strategy. Thus, in such games one could equivalently describe any result either in terms of underlying strategies and their complexity ($comp(\cdot)$) or in terms of machines and the number of states in them.

We shall establish below (by defining machines appropriately) that this equivalence between the two representations also holds in the negotiation game. Our approach to complexity will then be facilitated in machine terms as this will provide a more economical way to present the analysis. Each player’s strategy space in the negotiation game will be taken as the set of all machines and the players simultaneously and independently choose a single machine at the beginning of the negotiation game. This is the “machine game”, a term which we shall use interchangeably with the negotiation game.

Here, the extensive form of the stage game allows for many different machine specifications to equivalently represent a strategy.⁷ We adopt a particular machine specification that consists of two “sub-machines”, representing the two roles of each player.⁸

Definition 2 For each player i , a machine, $M_i = \{M_{ip}, M_{ir}\}$, consists of two sub-machines $M_{ip} = (Q_{ip}, q_{ip}^1, \lambda_{ip}, \mu_{ip})$ and $M_{ir} = (Q_{ir}, q_{ir}^1, \lambda_{ir}, \mu_{ir})$ where, for any $k, l = p, r$,

$$\begin{aligned} &Q_{ik} \text{ is the set of states; } q_{ik}^1 \text{ is the initial state belonging to } Q_{ik}; \\ &\lambda_{ik} : Q_{ik} \times D_{ik} \rightarrow C_i \text{ is the output function such that} \\ &\quad \lambda_{ik}(q_{ik}, d) \in C_{ik}(d) \quad \forall q_{ik} \in Q_{ik} \text{ and } \forall d \in D_{ik}; \text{ and} \\ &\mu_{ik} : Q_{ik} \times E \rightarrow Q_{il} \text{ is the transition function.} \end{aligned}$$

Let Φ_i denote the set of player i ’s machines in the machine game. Also, let Φ_i^t denote the set of player i ’s machines in the machine game starting with role distribution given in period t . Thus, if t is odd, $\Phi_i^t = \Phi_i$.

Each sub-machine in the above definition of a machine consists of a set of *distinct* states, an initial state and an output function enabling a player to play a given role. Transitions take place at the end of each period from a state in one sub-machine to a state in the other sub-machine as the roles are reversed. We also assume that each sub-machine has to have at least one state,⁹ and it shuts off if an agreement is reached.

⁷See also [16], [6], [7] and [19].

⁸There is no loss of generality in adopting any of the possible specifications in terms of implementing strategies. However, as we discuss later, the counting-the-number-of-states (of a machine) measure of complexity corresponds to different notions of complexity, in terms of underlying strategies, for different machine specifications.

⁹Note that the initial state of the sub-machine that operates in the second period is in fact redundant because the first state used by this sub-machine depends on the transition taking place between the first two periods of the game. We endow both sub-machines with an initial state for expositional ease.

Next, note that machines and strategies are equivalent in our setup. On the one hand, any machine implements a strategy in the negotiation game. Formally, any machine $M_i = \{M_{ip}, M_{ir}\} \in \Phi_i$, where $M_{ik} = (Q_{ik}, q_{ik}^1, \lambda_{ik}, \mu_{ik})$ for any $k = p, r$, implements (the same outcome at every subgame as) strategy f_i defined by

$$f_i(h, d) = \lambda_{ik}(q_i(h), d) \text{ for any } k, \text{ any } h \in H_{ik}^\infty \text{ and any } d \in D_{ik},$$

where $q_i(h) \in Q_{ik}$ denotes the state of M_i at history h .¹⁰ On the other hand, any strategy f_i can trivially be implemented by a machine. For example, consider a machine $M_i = \{M_{ip}, M_{ir}\}$ where, for any $k = p, r$, $M_{ik} = (Q_{ik}, q_{ik}^1, \lambda_{ik}, \mu_{ik})$, $Q_{ik} = H_{ik}^\infty$, $\lambda_{ik}(h, d) = f_i(h, d)$ and $\mu_{ik}(h, e) = (h, e)$ for any $h \in H_{ik}^\infty$, any $d \in D_{ik}$ and any $e \in E$.

We impose *no* restriction on the set of machines/strategies; each sub-machine may have any arbitrary (possibly infinite) number of states. This contrasts with Abreu and Rubinstein [1] and others who consider only finite automata. Assuming that machines have a finite number of states is itself a restriction on the players' choice of strategies.

Let $\|M_i\| = \sum_k |Q_{ik}|$ be the total number of states (or size) of machine M_i . The next result establishes that, for any strategy f_i , $\text{comp}(f_i)$ is equivalent to the total number of states of the smallest machine that implements f_i . It must be stressed here that the exact specification of a machine is important in qualifying this statement. In fact, it is precisely to establish this equivalence that we have chosen the above machine specification.¹¹

Proposition 1 *Fix any $f_i \in F_i$. Also, let $\overline{M}_i = \{\overline{M}_{ip}, \overline{M}_{ir}\}$ be a smallest machine that implements f_i ; that is, \overline{M}_i implements f_i and $\|\overline{M}_i\| \leq \|M'_i\|$ for any M'_i that implements f_i . Then, we have $|F_{ik}(f_i)| = \|\overline{M}_{ik}\|$ for any $k = p, r$ and thus $\|\overline{M}_i\| = \text{comp}(f_i)$.*

The proof extends Theorem 1 in KS and is omitted; see our working paper [13].

Given the above result, we now formally define the notion of complexity in terms of machines, as in Rubinstein [18] and Abreu and Rubinstein [1]: a machine M'_i is more complex than machine M_i if $\|M'_i\| > \|M_i\|$.¹²

We say that a machine is *minimal* if and only if each of its sub-machines has exactly one state. Such a machine implements the same actions in every period regardless of the

¹⁰Formally, if $h = (e^1, \dots, e^{t-1})$ then $q_i(h) = q_i^t$ where, for any $\tau \leq t$, q_i^τ is defined inductively as follows: $q_i^1 = q_{ip}^1$ (q_{ir}^1) if i is the proposer (responder) at $t = 1$, and for any $\tau > 1$, $q_i^\tau = \mu_{ik}(q_i^{\tau-1}, e^{\tau-1})$ if i 's role is k at $\tau - 1$.

¹¹Since in defining $\text{comp}(f_i)$ we consider continuation strategies *at the beginning of each period*, we need transitions between the states of a machine to take place *between periods* (in accordance with the continuation points chosen) and that each sub-machine has a distinct set of states (consistent with the continuation strategies in different roles being distinct).

¹²Binmore, Piccione, and Samuelson [3] propose another notion of complexity similar to the one considered in this paper and others. According to their "collapsing state condition", a machine M_i is less complex than another machine M'_i if the same implementation can be obtained by consolidating a collection of states in M'_i into a single state in M_i . Our results also hold under this notion of complexity.

history of the preceding periods, provided that the partial history within the current period is the same. Thus, it corresponds to a stationary strategy as in Definition 1. We shall henceforth use the terms “minimal” and “stationary” interchangeably.

To conclude the description of the machine game, we introduce some further notation. Let $M = (M_1, M_2)$ be a machine profile. Then, if M is the chosen machine profile, $T(M) \in \{1, 2, \dots, \infty\}$ refers to the period at which the negotiation game ends; $z(M) \in \Delta^2$ is the agreement offer if $T(M) < \infty$; $a^t(M) \in A$ is the disagreement game outcome in period t ; and $q_i^t(M)$ is the state of the active sub-machine of machine M_i in period t . Also, $\pi_i^t(M)$ denotes player i 's (discounted) average *continuation payoff* at period $t \leq T(M)$ if M is chosen. Thus, if the players reach an agreement on the partition $z = (z_1, z_2)$ at $T(M) < \infty$, $\pi_i^t(M) = (1 - \delta) \sum_{s=1}^{T-1} \delta^{s-1} u_i(a^s(M)) + \delta^{T-1} z_i$, and if an agreement is never reached, $\pi_i^t(M) = (1 - \delta) \sum_{s=1}^{\infty} \delta^{s-1} u_i(a^s(M))$. Also, with slight abuse of notation, let $\pi_i(M) = \pi_i^1(M)$. Finally, when the meaning is clear, we shall drop the argument M in the above terms and refer to them by T , z , a^t , q_i^t and π_i^t .

We now introduce an equilibrium notion that captures the players' preference for less complex machines. There are several ways of refining Nash equilibrium with complexity. For the main analysis we choose an equilibrium notion in which complexity cost enters each player's preferences *lexicographically*. In Section 4 below, we discuss how sharper results can be obtained when there is a positive tradeoff between complexity and payoff.

Definition 3 *A machine profile $M^* = (M_1^*, M_2^*)$ constitutes a Nash equilibrium of the machine game (NEM) if, for all i , (i) $\pi_i(M_i^*, M_{-i}^*) \geq \pi_i(M_i', M_{-i}^*) \forall M_i' \in \Phi_i$; (ii) there exists no $M_i' \in \Phi_i$ such that $\pi_i(M_i', M_{-i}^*) = \pi_i(M_i^*, M_{-i}^*)$ and $\|M_i^*\| > \|M_i'\|$.*

Note that since any strategy can be implemented by a machine it follows, by definition, that any NEM profile is a Nash equilibrium of the negotiation game.

NEM strategy profiles are not necessarily credible, however. A direct, and simple, way of introducing credibility is to consider NEM profiles that are subgame-perfect equilibria of the negotiation game without complexity cost.¹³

Definition 4 *A machine profile $M^* = (M_1^*, M_2^*)$ constitutes a subgame-perfect equilibrium of the machine game (SPEM) if M^* is both a NEM and a SPE.*

Given Proposition 1, we can equivalently define the above equilibrium concepts in terms of underlying strategies and the corresponding measure of complexity, $comp(\cdot)$. As mentioned earlier, we prefer the machine game analysis for its expositional economy.

¹³An alternative way to do this is to introduce trembles into the model and consider the limit of extensive form trembling hand equilibrium (Nash equilibrium with independent trembles at each information set) with complexity cost as the trembles become small. See Chatterjee and Sabourian [6][7].

2.3 Complexity and Efficiency

2.3.1 Some Preliminary Results

We begin by laying out two obvious, yet very important, properties of a NEM that will pave the way for the main results below. First, in any NEM, if the machine for some player is not minimal then every state of the machine must appear on the equilibrium path. Otherwise, the unused state can be “dropped” to reduce complexity cost without affecting the outcome and payoff, thereby contradicting the NEM assumption. This argument leads to the following Lemma.

Lemma 1 *Suppose that $M^* = (M_1^*, M_2^*)$ is a NEM. Then, (i) if $T(M^*) \geq 2$, every state in each player’s machine appears on the equilibrium path; (ii) if $T(M^*) \leq 2$, M_1^* and M_2^* are minimal.¹⁴*

Second, since any NEM profile $M^* = (M_1^*, M_2^*)$ is a Nash equilibrium of the negotiation game, we have $\pi_i(M_i^*, M_j^*) = \max_{f_i \in F_i} \pi_i(f_i, M_j^*)$ for any i and j , where, with some abuse of notation, $\pi_i(f_i, M_j^*)$ refers to i ’s payoff in the negotiation game where i and j play according to f_i and M_j^* respectively. More generally, NEM machines must be best responses (in terms of payoffs) *along* the equilibrium path of the negotiation game. Formally, for any $M = (M_i, M_j)$ and $q_j^\tau \equiv q_j^\tau(M)$, let $M_j(q_j^\tau) \in \Phi_j^\tau$ be the machine identical to M_j except it starts with the sub-machine operating in period τ with initial state q_j^τ . Also, with some abuse of notation, let $\pi_i(f_i, M_j^*(q_j^\tau))$ be i ’s payoff in the negotiation game that starts with role distribution as in period τ and is played by i and j according to $f_i \in F_i^\tau$ and $M_j^*(q_j^\tau)$, respectively. We then have the following.

Lemma 2 *Suppose that $M^* = (M_1^*, M_2^*)$ is a NEM. Then, $\forall i, j$ and $\forall \tau \leq T(M^*)$, $\pi_i^\tau(M^*) = \max_{f_i \in F_i^\tau} \pi_i(f_i, M_j^*(q_j^\tau))$.*

Lemmas 1 and 2 above are simple adaptations of some familiar arguments in the existing literature to our setup; we refer the reader to our working paper [13] for proofs.

2.3.2 Agreement, Stationarity and Efficiency

We first establish the following critical Lemma: if a NEM induces an agreement in a finite period then every state of the equilibrium machines occurs only once on the equilibrium path. Some of the steps of the proof use the arguments behind Lemmas 3 and 4 in Abreu and Rubinstein [1] which deliver their “tracking states” result. In that paper, machines are assumed to be finite, so any equilibrium must induce an outcome path that repeats perpetually, or “cycles”. Here, we do not impose finiteness of a machine; moreover, when there is an agreement at some finite period cycles clearly cannot happen.

¹⁴Note that although a player can choose a machine of any size it follows from Lemma 1 that, for any NEM profile $M^* = (M_1^*, M_2^*)$, M_i^* ($i = 1, 2$) must have a *countable* number of states.

Lemma 3 *Suppose $M^* = (M_1^*, M_2^*)$ is a NEM such that $T(M^*) < \infty$. Then, $q_i^t(M^*) \neq q_i^{t'}(M^*)$ for any $t, t' \leq T(M^*)$ and any i .*

Next, we present our first major result: if an agreement occurs at some finite date as a NEM outcome, then it must occur within the first stage (two periods) of the negotiation game, and thus, the associated machines (strategies) must be minimal (stationary).

Proposition 2 *Suppose $M^* = (M_1^*, M_2^*)$ is a NEM such that $T(M^*) < \infty$. Then, (i) $T(M^*) \leq 2$ and (ii) M_1^* and M_2^* are minimal, and hence, M^* is stationary.*

The first part of this claim holds because if a NEM induces an agreement beyond the first stage, then one of the players must be able to drop the last period's state of his machine without affecting the outcome of the game. The intuition for this is as follows. We know from Lemma 3 that the state of each player's machine occurring in the last period is distinct. This implies that (i) in the case where the final offer on the equilibrium path was proposed before by the same player, the proposer could reduce complexity cost simply by using one state to make the offer twice; and (ii) if the final offer occurs only once then the responder in the last period could reduce complexity cost by replacing his last period's state with any other state in his (sub-)machine and revising the corresponding output function to accept the final offer. Note that the argument in the latter case is of a different kind and relies critically on the fact that the output of a responder's sub-machine in a given state can be a function of the proposal.¹⁵

It immediately follows from Proposition 2 that any NEM involving an agreement must be efficient unless the agreement is in period 2. Therefore, any such NEM is almost efficient for a sufficiently large discount factor (efficient in the limit as $\delta \rightarrow 1$).

Corollary 1 *For any $\epsilon > 0$, there exists $\bar{\delta} < 1$ such that, for any $\delta \in (\bar{\delta}, 1)$, if M^* is a NEM with discount factor δ such that $T(M^*) < \infty$ then $\sum_i \pi_i(M^*) > 1 - \epsilon$.*

2.3.3 Perpetual Disagreement and Long-Run Efficiency

Let Ω^δ denote the set of SPEMs in the negotiation game with discount factor δ . We next show that, for sufficiently large δ , any SPEM outcome with perpetual disagreement must be at least *long-run* almost efficient; that is, if agreement never occurs then the players must eventually reach a finite period at which the sum of their continuation payoffs is approximately equal to one.

¹⁵Chatterjee and Sabourian [6][7] use similar intuition to derive a comparable result in a multi-player bargaining game. However, the details of the arguments are somewhat different because first we have to consider what happens in the disagreement game (after a rejection), and second, their analysis, in particular that in [7], is based on a different notion of complexity from one considered here.

Proposition 3 For any $\epsilon \in (0, 1)$, there exists $\bar{\delta} < 1$ such that, for any $\delta \in (\bar{\delta}, 1)$ and any $M^* \in \Omega^\delta$ with $T(M^*) = \infty$, there exists $\tau < \infty$ such that $\sum_i \pi_i^\tau(M^*) > 1 - \epsilon$.

The proof of Proposition 3 requires an interaction between complexity and perfection arguments. First, if some player i deviates from a SPEM by making another offer in some period t and the responder j rejects it then j 's continuation payoff at the next period must occur on the equilibrium path. The reason is as follows. By the complexity argument of Lemma 1, the state of i 's machine next period at $t + 1$ after j 's rejection at t must appear on the equilibrium path at some date t' ; but then, by Lemma 2, j 's corresponding continuation payoff at $t + 1$ must equal his equilibrium payoff at t' . Notice, however, that the complexity argument does not impose any restriction on the disagreement game outcome immediately after the deviating offer/rejection in t ; non-equilibrium outcomes can occur in this disagreement game. This is because our machine specification fixes the state of each player's (sub-)machine during each period (not at each decision node) and the machines can condition behavior in the disagreement game on the precise deviation offer.

Second, consider any period, call it $\tau + 1$, in which player j obtains his maximum continuation payoff in the proposer role; then we can show that, for any δ , the degree of inefficiency in the continuation payoffs at the *preceding* period τ cannot exceed $(1 - \delta)\beta_j$, where $\beta_j \equiv \sup_{a, a' \in A} [u_j(a) - u_j(a')]$. Otherwise, bargaining can be used by the other player i to break up the on-going disagreement at τ . This is because, by the previous complexity argument, the responder j in that period, who will be proposing next, cannot obtain more as of the next period $\tau + 1$ from rejecting the offer than what he is already getting from the original outcome. The upper bound $(1 - \delta)\beta_j$ on the degree of inefficiency is set by the importance of the current period in which the deviation can be followed by an off-the-equilibrium play of the disagreement game. As δ goes to 1, the inefficiency disappears at τ .

Proposition 3 does not however rule out the possibility of inefficiency early on in the negotiation game. It implies that, for any ϵ , any δ sufficiently high and any $M^* \in \Omega^\delta$ with $T(M^*) = \infty$, the game must reach some finite date τ such that the equilibrium payoffs satisfy

$$\sum_i \pi_i(M^*) \geq (1 - \delta) \sum_i \sum_{t=1}^{\tau-1} \delta^{t-1} u_i(a^t(M^*)) + \delta^{\tau-1}(1 - \epsilon) .$$

If τ was bounded then in the limit, as $\delta \rightarrow 1$ and $\epsilon \rightarrow 0$, the right-hand side of this inequality would be 1. However, we cannot guarantee such an ex ante efficiency result because τ may become arbitrarily large as $\delta \rightarrow 1$ and $\epsilon \rightarrow 0$. In particular, recall that τ is defined for each machine profile such that one of the players' continuation payoff at

$\tau + 1$ is at its maximum. Thus, τ depends on the the identity of the equilibrium machines and hence on the discount factor.^{16,17}

2.3.4 Main Characterization Results for SPEM

Now, we put together Propositions 2-3 and Corollary 1 to summarize the main properties of a SPEM. Every SPEM (in fact any NEM) inducing an agreement must do so in the very first stage of the negotiation game and hence be stationary, and with sufficiently patient players, every such SPEM is also almost efficient as is every SPEM inducing perpetual disagreement in the long-run.

Theorem 1 1. For any δ and any $M^* \in \Omega^\delta$, if $T(M^*) < \infty$ then $T(M^*) \leq 2$ and M^* is stationary.

2. For any $\epsilon \in (0, 1)$, there exists $\bar{\delta} < 1$ such that, for any $\delta \in (\bar{\delta}, 1)$ and any $M^* \in \Omega^\delta$,
- (a) if $T(M^*) < \infty$ then $\sum_i \pi_i(M^*) > 1 - \epsilon$;
 - (b) if $T(M^*) = \infty$ then there exists $\tau < \infty$ such that $\sum_i \pi_i^\tau(M^*) > 1 - \epsilon$.

It immediately follows from part 2(b) of the above Theorem that, with sufficiently high δ , if there exists no efficient action profile in G then the players cannot disagree forever; thus, by part 1, every SPEM must induce an agreement in the first stage, and hence, be stationary and almost efficient.

Corollary 2 Suppose that $\sum_i u_i(a) < 1$ for every $a \in A$. Then, for any $\epsilon \in (0, 1)$, there exists $\bar{\delta} < 1$ such that, for any $\delta \in (\bar{\delta}, 1)$, every $M^* \in \Omega^\delta$ is such that $T(M^*) \leq 2$, and hence, is stationary and $\sum_i \pi_i(M^*) > 1 - \epsilon$.

Any SPEM that induces an agreement is a stationary SPE. Also, since any stationary strategy can be implemented by a minimal machine, every stationary SPE is a SPEM. By considering the structure of stationary SPEs, we next characterize further properties of a SPEM. In particular, we (i) discuss the existence of a SPEM and (ii) identify, in the case of a SPEM *with an agreement*, a maximum bound on the degree of inefficiency for arbitrary δ , and consequently, obtain a set of sufficient conditions for full efficiency (part 2(a) of Theorem 1 is only an approximate efficiency result for sufficiently high δ).

¹⁶The critical period τ may also depend on ϵ if the equilibrium machines have infinite states and the maximum continuation payoffs are not defined. In such a case, τ has to be defined such that the relevant continuation payoff at period $\tau + 1$ approximates its supremum (by an amount less than ϵ). Thus, as $\epsilon \rightarrow 0$, τ may become unbounded.

¹⁷If we assume finite machines, any machine profile must generate cycles. Although such an assumption ensures that the maximum continuation payoffs are defined, it is not enough to guarantee that Proposition 3 implies ex ante efficiency in the limit as $\delta \rightarrow 1$. For this, we need for instance to assume additionally that the size of a machine is uniformly bounded (for any δ) so that the first cycle cannot continue beyond a fixed period.

Notice that for a pair of stationary strategies to constitute a SPE of the negotiation game only a Nash equilibrium of the disagreement game can be played after a rejection (on- or off-the-equilibrium path); otherwise, there will be a profitable deviation for some player as continuation payoffs are history-independent at the beginning of next period. This property of a stationary SPE has the following implications.

First, by Corollary 1 of BW, it implies that there exists a stationary SPE, and hence a (stationary) SPEM, if and only if the set of Nash equilibria in G is non-empty.

Second, by Proposition 1 of BW, a SPE payoff profile is unique and efficient if the action profile in the disagreement game is fixed and independent of the past for each period. It then follows from the above property that a stationary SPE payoff profile is unique and efficient if G has a unique Nash equilibrium. But, since our notion of a stationary strategy (Definition 1) allows for actions conditional on partial history within a period, a stationary SPE is not necessarily efficient if G has multiple Nash equilibria. In such cases, delays in agreement (even perpetual disagreement) and inefficiency can be sustained in a stationary SPE because the (Nash equilibrium) play of the disagreement game can be conditioned on the partial history, and therefore, a player who makes a deviating offer can be credibly punished in the disagreement game of the same period. (See our working paper [13] for an example of stationary SPE with perpetual disagreement.) However, the degree of punishment, and hence inefficiency, is bounded by the one-period payoff differences in the Nash equilibria of the disagreement game. Denote the set of (pure strategy) Nash equilibria of G by A^* , and let $b \equiv \max_i \sup_{a, a' \in A^*} [u_i(a) - u_i(a')]$; a formal statement of this argument is as follows.

Lemma 4 *If f is a stationary SPE of the negotiation game, $\sum_i \pi_i(f) \geq 1 - (1 - \delta)b$.*

Using this, we establish some cases in which a stationary SPE is efficient for any δ .

Lemma 5 *Suppose that one of the following conditions holds: (i) G has a unique Nash equilibrium, (ii) all Nash equilibria in G are strictly Pareto-ranked (i.e. $\forall a, a' \in A^*$, if $u_i(a) > u_i(a')$ then $u_j(a) > u_j(a')$), and (iii) $\sum_i u_i(a) < 1 - b \forall a \in A^*$. Then, for any δ , if f is a stationary SPE of the negotiation game, $\sum_i \pi_i(f) = 1$.*

The above two Lemmas immediately enable us to state the following.

Proposition 4 *Fix any δ and any $M^* \in \Omega^\delta$ such that $T(M^*) < \infty$. Then, $\sum_i \pi_i(M^*) \geq 1 - (1 - \delta)b$; moreover, if one of the conditions in Lemma 5 above holds, $\sum_i \pi_i(M^*) = 1$.*

We can also combine this Proposition with Corollary 2 to state the following.

Corollary 3 *Suppose that $\sum_i u_i(a) < 1 \forall a \in A$, and also that one of the conditions in Lemma 5 above holds. Then, there exists $\bar{\delta} < 1$ such that, for any $\delta \in (\bar{\delta}, 1)$, every $M^* \in \Omega^\delta$ is such that $\sum_i \pi_i(M^*) = 1$.*

3 The Role of Transaction Costs

Let us now investigate the impact of transaction costs on our selection results in the negotiation game. Consider the following “costly negotiation game”.

As in AF, we assume that at the beginning of every period each player must pay a participation/transaction cost $\rho \in (0, \frac{1}{2}]$ to enter bargaining, but *not* the disagreement game. There are several ways to think about how this decision is made. The players can pay the cost either simultaneously or sequentially. This is immaterial to AF’s original analysis, but the extensive form matters for the precise details of the results here in the negotiation game. In order to stay synchronized with the sequential structure of the bargaining, we assume that at each period t the proposer first decides whether or not to pay ρ . If he makes the payment, the responder at t then decides whether or not to pay ρ . Bargaining in that period occurs if and only if both players sink the cost; otherwise the players move directly to the disagreement game before reaching the next period. (We discuss the simultaneous participation case in the next section.)

Let us now modify some notation for the costly negotiation game. Let \mathcal{I} and \mathcal{N} denote a player’s decision to pay (participate) and not pay ρ , respectively. For each i , the set of partial histories within a period for a role is now one of the following

$$\begin{aligned} D_{ip} &= \{\emptyset, (\mathcal{I}, \mathcal{I}), (\mathcal{N}), (\mathcal{I}, \mathcal{N}), (\mathcal{I}, \mathcal{I}, x^i, N)\}_{x^i \in \Delta^2} \\ D_{ir} &= \{(\mathcal{I}), (\mathcal{I}, \mathcal{I}, x^j), (\mathcal{N}), (\mathcal{I}, \mathcal{N}), (\mathcal{I}, \mathcal{I}, x^j, N)\}_{x^j \in \Delta^2}, \end{aligned}$$

where, for example, $(\mathcal{I}, \mathcal{I})$ represents the partial history of sequential payment of the cost by both players. We similarly modify the definition of E , the set of outcomes in a period. The set of player i ’s actions is now given by $C_i \equiv \mathcal{I} \cup \mathcal{N} \cup \Delta^2 \cup Y \cup N \cup A_i$, and

$$C_{ik}(d) = \begin{cases} \{\mathcal{I}, \mathcal{N}\} & \text{if } k = p \text{ and } d = \emptyset \text{ or if } k = r \text{ and } d = (\mathcal{I}) \\ \Delta^2 & \text{if } k = p \text{ and } d = (\mathcal{I}, \mathcal{I}) \\ \{Y, N\} & \text{if } k = r \text{ and } d \in \{(\mathcal{I}, \mathcal{I}, x^j)\}_{x^j \in \Delta^2} \\ A_i & \text{if } k = p, r \text{ and } d \in \{(\mathcal{N}), (\mathcal{I}, \mathcal{N}), (\mathcal{I}, \mathcal{I}, x^i, N)\}_{x^i \in \Delta^2} \end{cases}.$$

With this set of modifications, all previous notation/definitions on histories, strategies, complexity and machines also carry to the costly negotiation game.

Next, we define the payoffs. Let ρ_i^t be the (discounted) sum of participation costs that player i incurs between period t and the end of the game T under profile M . Then, we can define player i ’s (discounted) average continuation payoff at t (given M) as

$$\pi_i^t = \begin{cases} (1 - \delta) \left[\sum_{\tau=t}^{\infty} \delta^{\tau-t} u_i(a^\tau) - \rho_i^t \right] & \text{if } T = \infty \\ (1 - \delta) \left[\sum_{\tau=t}^{T-1} \delta^{\tau-t} u_i(a^\tau) - \rho_i^t \right] + \delta^{T-t} z_i & \text{if } t < T < \infty \\ z_i - (1 - \delta)\rho & \text{if } t = T < \infty. \end{cases}$$

Before characterizing the set of SPEMs in this game, note that BW's Folk theorem type characterization of the set of SPE payoffs in the negotiation game without transaction costs extends to one with transaction costs. In fact, here one can establish a full Folk theorem. With transaction costs and sufficiently patient players, not only there are multiple SPE outcome paths involving immediate agreement, delays in agreement and perpetual disagreement, one can also show that any feasible and individually rational payoffs (obtainable by pure strategies) can be sustained in a SPE.¹⁸

Let us now investigate the set of NEMs/SPEMs in the costly negotiation game. First, note, by exactly the same reasoning as in the previous section, that the properties of NEM profiles specified in Lemmas 1-3 above also hold for the negotiation game with transaction costs. To save space we shall omit their statements and proofs. Next, we establish, by appealing to Lemmas 1-3 (as in Proposition 2 above), that if an agreement takes place in some finite period as a NEM outcome of the costly negotiation game then it must occur in the first stage.

Proposition 5 *Fix any $\rho > 0$, and suppose that M^* is a NEM in the costly negotiation game such that $T(M^*) < \infty$. Then, (i) $T(M^*) \leq 2$ and (ii) M^* is stationary.*

As in the negotiation model without transaction costs, it follows that any SPEM inducing an agreement here is stationary, and therefore, we need to characterize the set of stationary SPEs in the costly negotiation game. As before, let A^* be the set of Nash equilibria in G and $b \equiv \max_i \sup_{a, a' \in A^*} [u_i(a) - u_i(a')]$.

Clearly, any stationary SPE in the costly negotiation game induces either an agreement in the first stage or perpetual disagreement. However, if $\rho > b$, the latter is the only stationary SPE outcome; otherwise, it would be better for the responder at the agreement date to hold up the payment of the participation cost. The reasoning is similar to that in AF which shows that perpetual disagreement is the unique stationary SPE outcome in the Rubinstein model with transaction costs; see Section 4 below.

Lemma 6 *If $\rho > b$, every stationary SPE of the costly negotiation game induces perpetual disagreement.*

Putting together Proposition 5 and Lemma 6, we immediately see that, if $\rho > b$, there cannot be a SPEM inducing an agreement because in this case perpetual disagreement is the unique stationary SPE outcome. This argument leads to our next Theorem.

Theorem 2 *Every SPEM profile M^* in the costly negotiation game is such that:*

1. *if $\rho > b$ then $T(M^*) = \infty$;*
2. *if $0 < \rho \leq b$ then either $T(M^*) \leq 2$ or $T(M^*) = \infty$.*

¹⁸To save space we shall not provide a proof of this claim. However, note that with transaction costs the multiplicity problem is worse because of the coordination issue arising from the fact that bargaining can occur if and only if *both* players decide to participate.

The following Corollaries further highlight the dramatic impact of transaction costs in the negotiation game.

Corollary 4 *Suppose that $\sum_i u_i(a) < 1$ for every $a \in A$. Then, for any $\rho > b$, every SPEM in the costly negotiation game is inefficient.*

Corollary 5 *Suppose that G has a unique Nash equilibrium. Then, for any $\rho > 0$, every SPEM profile M^* in the costly negotiation game is such that $T(M^*) = \infty$, and if in addition $\sum_i u_i(a) < 1 \forall a \in A$, it is inefficient.*

Notice the stark contrast between these two Corollaries and Corollaries 2 and 3 above for the case in which every disagreement game outcome is inefficient ($\sum_i u_i(a) < 1 \forall a \in A$). In particular, if G has a unique Nash equilibrium, the only feasible SPEM outcome in the negotiation game without transaction costs in this case is an agreement in the first stage which is therefore efficient (in the limit); with any positive transaction cost, we only have perpetual disagreement and *inefficiency*.

4 Extensions and Additional Results

We have thus far analyzed a particular setup - extensive form, complexity/machine specifications and equilibrium concepts - to address our questions concerning the Coase Theorem. In this section, we discuss how our central message can in fact be made sharper by considering alternative setups.

4.1 Extending the Complexity Analysis in the Negotiation Game without and with Transaction Costs

In this sub-section, we consider informally three extensions of the analysis of the negotiation game without and with transaction costs. To save space we shall not provide a formal analysis here; see our working papers [13][14] for detailed statements and proofs.

Positive Complexity Cost. In this paper we have considered lexicographic preferences in which complexity cost matters only after the payoff. Our results also hold when there is *any* positive tradeoff between payoff and complexity in each player's preferences. In fact, with such a tradeoff the set of equilibria can be limited further and a stronger set of selection results are obtained.

First, we can show that in this case if the players are sufficiently patient then every SPEM in the negotiation game without transaction costs must be stationary, and hence, almost efficient (not just in the long-run or when there is an agreement).¹⁹

¹⁹Almost efficiency is replaced by efficiency if G satisfies one of the conditions in Lemma 5.

Second, in the costly negotiation game we have assumed that the players sequentially decide whether to pay ρ and participate in the bargaining. It turns out that, when complexity is treated lexicographically, the selection results in Section 3 do depend on this sequential participation assumption. If we instead assume simultaneous participation, to obtain the selection results, there is the additional coordination issue on the participation decisions of the two players. But, if there is any positive tradeoff between payoff and complexity in each player's preferences, we can establish the results of Section 3 on perpetual disagreement/inefficiency for the costly negotiation game even with simultaneous participation decisions, as long as the players are sufficiently patient.

Alternative Complexity/Machine Specifications. Since each stage game of the negotiation game has a sequential structure, we can adopt alternative machine specifications that employ more frequent transitions and hence account for finer partitions of histories, thereby corresponding to a different measure of complexity. For instance, in our working papers [13][14] we also consider a machine definition that maintains the role distinction and employs distinct sub-machines to play the bargaining and the disagreement game within each period. Such a machine has four sub-machines (two in each role) with transition occurring twice within each period, and its complexity (the number of states) is equivalent to counting the number of continuation strategies that the underlying strategy induces at the beginning and at the disagreement game of every period for each role. Using this machine definition, we derive a set of NEM/SPEM results containing much the same flavor as the corresponding results in this paper but are sharper: the results in the negotiation game without and with transaction costs do not depend on the discount factor and the transaction cost, respectively.

More specifically, recall that many of the efficiency results without transaction costs above depend on a sufficiently large δ because a deviating offer can be met by an off-the-equilibrium response in the disagreement game within the same period. For a similar reason, in the costly negotiation game perpetual disagreement is not the unique (stationary) SPEM outcome if ρ is smaller than what a player can lose in the disagreement game. However, with the four sub-machine specification described above, the disagreement outcomes no longer depend on the partial history within a period under a minimal machine profile. Therefore, with this machine definition, we can establish the following two results for *any* δ . First, without transaction costs, every SPEM either induces an agreement in the first stage and is efficient or it is long-run almost efficient.²⁰ Second, with *any* positive transaction costs every SPEM induces perpetual disagreement.

The No Discounting Case. Our results are not affected if there is no discounting. In fact, by taking $\delta = 1$ and using the Limit of the Mean criterion, a sharper result can be obtained: every SPEM in the negotiation game is efficient.

²⁰If in addition we have either $\sum_i u_i(a) < 1 \forall a \in A$ or a positive tradeoff between payoff and complexity in preferences, we can show every SPEM induces an immediate agreement and is efficient.

4.2 Complexity and Transaction Costs in the Rubinstein Bargaining Model

The complexity analysis thus far endorses the conclusion that the Coase Theorem is valid in the negotiation model if and only if there are no transaction costs. In fact, even sharper results bearing the same message arise from the Rubinstein [17] bargaining setup. We know that every SPE, and hence SPEM, of this game is efficient (and stationary); let us next consider the “costly bargaining game” of AF (Rubinstein model with transaction costs). Unless stated otherwise, the same notation is used as before.

Two players engage in Rubinstein bargaining over a unit surplus (which accrues just once, not periodically) with player 1 being the proposer at odd dates and player 2 at even dates. There is no disagreement game to be played after a rejection. However, in order for the players to enter bargaining in each period both must sequentially pay a participation cost $\rho \in [0, \frac{1}{2}]$ at the beginning of each period. At each date the proposer first decides whether or not to pay ρ ; if he makes the payment, the responder then decides whether or not to pay ρ . Bargaining in that period occurs if and only if both players sink the cost; otherwise the players move directly onto the next period.²¹

With appropriate notational modifications (for example, $D_{ip} = \{\emptyset, (\mathcal{I}, \mathcal{I})\}$, $D_{ir} = \{(\mathcal{I}), (\mathcal{I}, \mathcal{I}, x^j)\}_{x^j \in \Delta^2}$ and $C_i = \mathcal{I} \cup \mathcal{N} \cup \Delta^2 \cup Y \cup N$), the definitions of histories, strategies, complexity and machines remain as before. Denote by $\rho_i(f)$ the (discounted) sum of transaction costs that player i pays along the entire outcome path under f . If f induces an agreement $z \in \Delta^2$ in a finite period T , the payoff to i is $\pi_i(f) = \delta^{T-1} z_i - \rho_i(f)$, while if f induces perpetual disagreement i 's payoff is $\pi_i(f) = -\rho_i(f)$.

AF show that the costly bargaining game admits a unique stationary SPE in which neither player ever pays ρ and therefore disagreement persists forever. Using this stationary SPE as a penal code, they then show that there exists a large number of non-stationary SPE outcomes as long as ρ is not too large.²²

Now, we show that, for any $\rho > 0$, the unique stationary SPE that induces perpetual disagreement is in fact the unique SPEM in the costly bargaining game. To demonstrate this, note first that Lemmas 1 and 2 in Section 2 also hold here in the costly bargaining game. To save space we shall omit their statements and proofs. Using these Lemmas, we next show that, in any SPEM profile M^* , if there is an agreement it must be at period 1, and hence, M^* must be stationary.

Proposition 6 *Fix any $\rho > 0$, and suppose that M^* is a SPEM in the costly bargaining game such that $T(M^*) < \infty$. Then, (i) $T(M^*) = 1$ and (ii) M^* is stationary.*

²¹Every result and its proof in this sub-section apply to the case of simultaneous participation decisions (AF state their main result assuming such an extensive form), and also to the no discounting case.

²²AF also report a stronger result in which perpetual disagreement is the unique SPE when, at each period, players forget the past history with a probability less than, but sufficiently close to, 1.

The basic arguments here are similar to those for Proposition 3 above. Because of complexity considerations (Lemmas 1 and 2) any deviation can be punished only by what happens on the equilibrium path. This implies that, if an agreement is reached beyond the first period and a deviating offer is made in the last period, subsequent rejection induces the same agreement with at least two periods of further delay. Thus, because of discounting and/or positive transaction cost, there must exist a deviating offer for the last period's proposer which the responder will accept and improves the former's payoff.

Since the unique stationary SPE induces perpetual disagreement, it follows from Proposition 6 that every SPEM profile M^* is such that $T(M^*) = \infty$ and each player receives a zero payoff; hence, neither player pays ρ in any period and each M_i^* is minimal.

Theorem 3 *Fix any $\rho > 0$. For any SPEM profile M^* in the costly bargaining game, we have (i) $T(M^*) = \infty$, (ii) neither player pays ρ in any period and (iii) M^* is stationary.*

Complexity together with any positive transaction cost therefore selects a unique equilibrium outcome in the Rubinstein bargaining model, which is extremely *inefficient*.

5 Appendix: Proofs

Proof of Lemma 3. Suppose not; then \exists a NEM profile $M^* = \{M_1^*, M_2^*\}$, where $M_{ik}^* = (Q_{ik}^*, q_{ik}^{1*}, \lambda_{ik}^*, \mu_{ik}^*)$ for each i and k , such that $T(M^*) < \infty$ and $q_i^v = q_i^{v'}$ for some i and two distinct dates $v, v' \leq T$. We now show there must be a cycle, which contradicts $T(M^*) < \infty$.

Define $\tau_i \equiv \min\{t \mid q_i^t = q_i^{t'} \text{ for some distinct dates } t \text{ and } t'\}$. Also, let τ_i' be the minimal $t > \tau_i$ such that $q_i^t = q_i^{\tau_i}$. We proceed in the following steps.

Step 0: $\forall i, j$ and $\forall t, t' \leq T(M^*)$, if $q_j^t(M^*) = q_j^{t'}(M^*)$ then $\pi_i^t(M^*) = \pi_i^{t'}(M^*)$.

This immediately follows from Lemma 2.

Step 1: $\exists t \leq T$ such that $t \neq \tau_i$ and $q_j^t = q_j^{\tau_i}$.

Suppose not. Then, consider player j using another machine $M_j' = \{M_{jp}', M_{jr}'\}$, where $M_{jk}' = (Q_{jk}', q_{jk}'^V, \lambda_{jk}', \mu_{jk}')$ for $k = p, r$, which is identical to M_j^* except that (i) $q_j^{\tau_i}$ is dropped (thus $Q_{jl}' = Q_{jl}^* \setminus q_j^{\tau_i}$) and (ii) the transition function is such that $\mu_{jl}'(q_j^{\tau_i-1}, e^{\tau_i-1}) = q_j^{\tau_i}$, where j plays role l at τ_i and l' at $\tau_i - 1$.

Since $q_j^{\tau_i}$ is distinct, playing M_j' against M_i^* generates the same outcome path as M_j^* up to period $\tau_i - 1$ followed by the outcome path between τ_i' and T (thereby making the agreement occur sooner). Notice that, since $\pi_j^{\tau_i} = \pi_j^{\tau_i'}$ from Step 0, $\pi_j(M_i^*, M_j') = \pi_j(M_i^*, M_j^*)$. Also, since $q_j^{\tau_i}$ is dropped, $\|M_j^*\| > \|M_j'\|$. This contradicts NEM.

Step 2: $\tau_i = \tau_j$, where $\tau_j = \min\{t \mid q_j^t = q_j^{t'} \text{ for some distinct dates } t \text{ and } t'\}$.

By Step 1, τ_j is well-defined and $\tau_j \leq \tau_i$. Also, we can use exactly the same reasoning as in Step 1 to show that $q_i^{\tau_j}$ is not distinct, and hence $\tau_j \geq \tau_i$. Therefore, $\tau_j = \tau_i$.

Step 3: $\tau'_i = \tau'_j$, where τ'_j is the minimal $t > \tau_j$ such that $q_j^t = q_j^{\tau'_j}$.

Suppose not and assume $\tau'_i > \tau'_j$. Let $\tau = \tau_i = \tau_j$. There are two cases to consider.

Case A: $q_j^{\tau'_i} \neq q_j^t \forall t < \tau'_i$.

Now, consider another machine $M'_j = \{M'_{jp}, M'_{jr}\}$ identical to M_j^* except that (i) $q_j^{\tau'_i}$ is dropped (thus $Q'_{jl} = Q_{jl}^* \setminus q_j^{\tau'_i}$) and (ii) the transition function is such that $\mu'_{jl'}(q_j^{\tau'_i-1}, e^{\tau'_i-1}) = q_j^\tau$, where j plays role l at τ'_i and l' at $\tau'_i - 1$.

Since $q_i^\tau = q_i^{\tau'_i}$ and $q_j^{\tau'_i}$ does not happen before τ'_i , playing M'_j against M_i^* generates the same outcome path up to $\tau'_i - 1$ and then replicates the outcome path between τ and $\tau'_i - 1$ *ad infinitum*. Thus, j 's corresponding continuation payoff at τ is:

$$\pi_j^\tau(M_i^*, M'_j) = \frac{1 - \delta}{1 - \delta^{\tau'_i - \tau}} \sum_{t=\tau}^{\tau'_i - 1} \delta^{t-\tau} u_j(a^t) = \pi_j^\tau(M_i^*, M_j^*),$$

where the last equality follows because Step 0 implies $\pi_j^{\tau'_i}(M_i^*, M_j^*) = \pi_j^\tau(M_i^*, M_j^*)$.

Since (M_i^*, M'_j) and M^* induce the same outcome before τ , it follows that $\pi_j(M_i^*, M'_j) = \pi_j(M^*)$. But then, since $q_j^{\tau'_i}$ is dropped, $\|M_i^*\| > \|M'_j\|$, which contradicts NEM.

Case B: $q_j^{\tau'_i} = q_j^s$ for some $\tau \leq s < \tau'_i$.

If $s = \tau$ then the outcome path between τ and $\tau'_i - 1$ repeats perpetually, contradicting $T < \infty$. So, it must be that $s > \tau$. In this case, consider i using another machine $M'_i = \{M'_{ip}, M'_{ir}\}$ which is identical to M_i^* except that (i) q_i^τ is dropped and (ii) $\mu'_{il'}(q_i^{\tau-1}, e^{\tau-1}) = q_i^{\tau'_j}$ and $\mu'_{il'}(q_i^{\tau'_i-1}, e^{\tau'_i-1}) = q_i^s$, where i plays role l at τ and l' at $\tau - 1$.

Playing M'_i against M_j^* generates the same outcome path up to $\tau - 1$, followed by the outcome path between τ'_j and $\tau'_i - 1$, and then, repeats the outcome path between s and $\tau'_i - 1$ perpetually. Thus, we have

$$\pi_i^\tau(M'_i, M_j^*) = (1 - \delta) \sum_{t=\tau'_j}^{\tau'_i - 1} \delta^{t-\tau'_j} u_i(a^t) + \frac{(1 - \delta) \delta^{\tau'_i - \tau'_j}}{1 - \delta^{\tau'_i - s}} \sum_{t=s}^{\tau'_i - 1} \delta^{t-s} u_i(a^t). \quad (1)$$

Now, since $q_j^\tau = q_j^{\tau'_j}$ and $q_i^{\tau'_i} = q_i^s$, Step 0 implies that $\pi_i^\tau(M^*) = \pi_i^{\tau'_j}(M^*)$ and $\pi_i^{\tau'_i} = \pi_i^s = (1 - \delta) \sum_{t=s}^{\tau'_i - 1} \delta^{t-s} u_i(a^t) + \delta^{\tau'_i - s} \pi_i^{\tau'_i}$. From this, we can then derive

$$\pi_i^\tau(M^*) = (1 - \delta) \sum_{t=\tau'_j}^{\tau'_i - 1} \delta^{t-\tau'_j} u_i(a^t) + \frac{(1 - \delta) \delta^{\tau'_i - \tau'_j}}{1 - \delta^{\tau'_i - s}} \sum_{t=s}^{\tau'_i - 1} \delta^{t-s} u_i(a^t). \quad (2)$$

By (1) and (2), $\pi_i^\tau(M'_i, M_j^*) = \pi_i^\tau(M^*)$. But, since $\|M_i^*\| > \|M'_i\|$, this contradicts NEM.

Steps 2 and 3 imply that M^* eventually induces a cyclical outcome path (the outcome path between τ and $\tau_i' - 1$ cycles perpetually); but this contradicts $T < \infty$. \parallel

Proof of Proposition 2. By Lemma 1, if part (i) of the claim is true, part (ii) must also be true. To show (i), suppose otherwise. Then, an agreement $z \in \Delta^2$ occurs at some finite $T > 2$ on the equilibrium path. Suppose player i is the proposer at T . There are two possible cases that can occur on the equilibrium path.

Case A: $x^\tau = z$ at some $\tau < T$ where i proposes.

Consider another machine $M_i' = \{M_{ip}', M_{ir}'\}$ which is identical to M_i^* except that (i) q_i^T is dropped and (ii) $\mu_{ir}'(q_i^{T-1}, e^{T-1}) = q_i^\tau$. Since $T > 2$ and, by Lemma 3, q_i^T appears for the first time at T on the original equilibrium path, M_i' is well-defined; furthermore, since $\lambda_{ip}'(q_i^\tau, \emptyset) = \lambda_{ip}^*(q_i^\tau, \emptyset) = z$, M_i' (given M_j^*) generates an identical outcome path and payoff as M_i^* . This contradicts NEM because $\|M_i^*\| > \|M_i'\|$ (q_i^T is dropped).

Case B: $x^\tau \neq z \forall \tau < T$ where i proposes.

Consider another machine $M_j' = \{M_{jp}', M_{jr}'\}$ which is identical to M_j^* except that (i) q_j^T is dropped, (ii) $\mu_{jp}'(q_j^{T-1}, e^{T-1}) = q_j \neq q_j^T$ for some arbitrary but fixed $q_j \in Q_{jr}'$ (such q_j exists since we have $T > 2$ and, by Lemma 3, q_j^T is distinct) and (iii) $\lambda_{jr}'(q_j, z) = Y$.

Since the offer z does not appear anywhere before T on the original equilibrium path when i proposes, M_j' (given M_i^*) generates an identical outcome path and payoff as M_j^* . This contradicts NEM because $\|M_j^*\| > \|M_j'\|$ (q_j^T is dropped). \parallel

Proof of Proposition 3. We proceed by first establishing the following claim.

Claim: Fix any $\eta > 0$, any δ and any $M^* \in \Omega^\delta$ with $T(M^*) = \infty$. Then, there exists $\tau \in \infty$ such that $\sum_i \pi_i^\tau(M^*) \geq 1 - (1 - \delta)\beta_2 - \eta$, where $\beta_2 \equiv \sup_{a, a' \in A} [u_2(a) - u_2(a')]$.

To prove this claim, first define $t_{ik} \equiv \{t \mid i \text{ plays role } k\}$ and $\tau_\eta \equiv \min\{t \in t_{2p} \mid \pi_2^t + \eta > \pi_2^{t'} \forall t' \in t_{2p}\}$. Next, given M_1^* , consider player 2's continuation payoff after rejecting any offer in any period belonging to t_{2r} . Since every state of M_1^* appears on the equilibrium path of M^* (Lemma 1) and $\pi_2^t = \max_{f_2 \in F_2^t} (f_2, M_1^*(q_1^t))$, $\forall t$ (Lemma 2), 2's continuation payoff at the next period if he rejects any offer (given M_1^*) is at most $\sup_{t \in t_{2p}} \pi_2^t$. We also have $\pi_2^{\tau_\eta} + \eta > \pi_2^t$, $\forall t \in t_{2p}$. Thus, since M^* is a SPE profile, machine M_2^* must always accept an offer $z' = (1 - \pi_{2r}^{\max}, \pi_{2r}^{\max}) \in \Delta^2$, where π_{2r}^{\max} is defined by

$$\pi_{2r}^{\max} \equiv (1 - \delta) \sup_{a \in A} u_2(a) + \delta (\pi_2^{\tau_\eta} + \eta) . \quad (3)$$

Now, consider another machine $M_1' = \{M_{1p}', M_{1r}'\}$ that is identical to M_1^* except that $\lambda_{1p}'(q_1^{\tau_\eta-1}, \emptyset) = z'$. Since M_2^* always accepts the offer z' and M_1' differs from M_1^* only in offers conditional on $q_1^{\tau_\eta-1}$, (M_1', M_2^*) results in agreement z' in period $\tau \equiv \min\{t \mid q_1^t = q_1^{\tau_\eta-1}\}$. But, since M^* is a SPEM, a deviation to M_1' is not profitable; thus,

$$\pi_1^\tau \geq 1 - \pi_{2r}^{\max} . \quad (4)$$

Also, since $q_1^\tau = q_1^{\tau\eta^{-1}}$, by Lemma 2 we have $\pi_2^\tau = \pi_2^{\tau\eta^{-1}}$. This implies that

$$\pi_2^\tau = (1 - \delta)u_2(a^{\tau\eta^{-1}}) + \delta\pi_2^{\tau\eta} . \quad (5)$$

Since $\sup_{a \in A} u_2(a) - u_2(a^{\tau\eta^{-1}}) \leq \beta_2$, we have, by (3) and (5), $\pi_{2r}^{\max} - \pi_2^\tau \leq (1 - \delta)\beta_2 + \delta\eta$. This implies that $1 - \pi_{2r}^{\max} \geq 1 - \pi_2^\tau - ((1 - \delta)\beta_2 + \eta)$. Then, together with (4), we obtain $\sum_i \pi_i^\tau \geq 1 - ((1 - \delta)\beta_2 + \eta)$, which proves the claim.

Now, fix any $\epsilon \in (0, 1)$ and any $\eta < \frac{\min\{\beta_2, \epsilon\}}{2}$. Let $\bar{\delta} = 1 - \frac{\eta}{\beta_2}$; clearly $\bar{\delta} \in (0, 1)$. Consider any $\delta \in (\bar{\delta}, 1)$ and any $M^* \in \Omega^\delta$ with $T(M^*) = \infty$. It then follows from the above claim that there exists τ such that $\sum_i \pi_i(M^*) \geq 1 - (1 - \delta)\beta_2 - \eta > 1 - \epsilon$, as in the Proposition. \parallel

Proof of Lemma 4. Fix any $\epsilon > 0$, and consider player 1 making a deviating offer $z' = (z'_1, z'_2) \in \Delta^2$ at $t = 1$ such that $z'_2 = \pi_2(f) + (1 - \delta)b + \epsilon$.

Since f is a stationary SPE, it follows from the definition of b that f_2 will accept z' at $t = 1$ (note that any disagreement outcome must be a Nash equilibrium of G). Also, such deviation must be unprofitable for player 1; therefore $\pi_1(f) \geq z'_1$. This, together with the definition of z'_2 , implies that $\pi_1(f) \geq 1 - \pi_2(f) - (1 - \delta)b - \epsilon$. Since the last inequality holds for any $\epsilon > 0$ it follows that $\sum_i \pi_i(f) \geq 1 - (1 - \delta)b$. \parallel

Proof of Lemma 5. (i) If G has a unique Nash equilibrium, $b = 0$; the claim follows from Lemma 4.

(ii) Suppose not; then there is a stationary SPE f such that $\sum_i \pi_i(f) < 1 - \epsilon$ for some $\epsilon > 0$ with no agreement at $t = 1$. Let $a \in A^*$ be the disagreement outcome induced by f in $t = 1$. Next, consider player 1 making a deviating offer $z' = (z'_1, z'_2) \in \Delta^2$ at $t = 1$ such that $z'_2 = \pi_2(f) + \epsilon$. Since $z'_1 = 1 - z'_2 > \pi_1(f)$ and f is a stationary SPE, f_2 must reject z' , immediately followed by some $a' \in A^*$ such that $u_2(a') > u_2(a)$ and $u_1(a') \leq u_1(a)$. But, this contradicts that all Nash equilibria in G are strictly Pareto-ranked.

(iii) Let f be a stationary SPE which is inefficient. Consider two possible cases.

Case A: *There is one period of delay, followed by an agreement.* We know from Lemma 4 that $\sum_i \pi_i(f) \geq 1 - (1 - \delta)b$. Since f induces a Nash equilibrium of G in any disagreement game, this implies that $(1 - \delta)\sum_i u_i(a) + \delta \geq 1 - (1 - \delta)b$ for some $a \in A^*$. But, this is clearly not possible if $\sum_i u_i(a) < 1 - b \forall a \in A^*$, as in the claim.

Case B: *There is infinite delay.* Let $a^1, a^2 \in A^*$ be the disagreement outcomes in periods 1 and 2, respectively. Then, by Lemma 4, we must have $\frac{\sum_i u_i(a^1) + \delta \sum_i u_i(a^2)}{1 + \delta} \geq 1 - (1 - \delta)b$. But, this is not possible if $\sum_i u_i(a) < 1 - b \forall a \in A^*$. \parallel

Proof of Proposition 5. If part (i) of the claim is true, part (ii) must be true because of Lemma 1. To show (i), suppose otherwise. Then, there exists a NEM profile

M^* in the costly negotiation game that induces some agreement $z \in \Delta^2$ at some finite $T > 2$. Let i be the proposer at T . Consider two possible cases.

Case A: Bargaining occurs at some $\tau < T$ where $\tau \in t_{ip} \equiv \{t \mid i \text{ plays role } p\}$.

But then, use the arguments in Proposition 2 to derive a contradiction against NEM.

Case B: No bargaining occurs at any $\tau < T$ where $\tau \in t_{ip}$.

Here, there are two sub-cases that can occur on the equilibrium path.

Case B1: i does not pay ρ at any $\tau < T$ where $\tau \in t_{ip}$.

Consider another machine $M'_j = \{M'_{jp}, M'_{jr}\}$ identical to M_j^* except (i) q_j^T is dropped, (ii) $\mu'_{jp}(q_j^{T-1}, e^{T-1}) = q_j \neq q_j^T$ for some arbitrary but fixed $q_j \in Q_{jr}$ (such q_j exists since $T > 2$ and q_j^T is distinct by Lemma 3), (iii) $\lambda'_{jr}(q_j, (\mathcal{I})) = \mathcal{I}$ and (iv) $\lambda'_{jr}(q_j, (\mathcal{I}, \mathcal{I}, z)) = Y$.

Since q_j^T is distinct and the partial histories (\mathcal{I}) and $(\mathcal{I}, \mathcal{I}, z)$ do not appear at any $\tau < T$, where $\tau \in t_{ip}$, on the original equilibrium path, M'_j (given M_i^*) does not affect the outcome and payoff. But, since q_j^T is dropped, $\|M_j^*\| > \|M'_j\|$. This contradicts NEM.

Case B2: $\exists \tau < T$ where $\tau \in t_{ip}$ such that i pays ρ and j does not.

Consider another machine $M'_i = \{M'_{ip}, M'_{ir}\}$ identical to M_i^* except (i) q_i^T is dropped, (ii) $\mu'_{ir}(q_i^{T-1}, e^{T-1}) = q_i^T$ (note that $q_i^T \neq q_i^T$ since $T > 2$ and q_i^T is distinct by Lemma 3), and (iii) $\lambda'_{ip}(q_i^T, (\mathcal{I}, \mathcal{I})) = z$.

Again, since q_j^T is distinct and j does not pay ρ at τ on the original equilibrium path, M'_i (given M_j^*) generates the same outcome path and payoff as M_i^* . But, since q_i^T is dropped, $\|M_i^*\| > \|M'_i\|$. This contradicts NEM. \parallel

Proof of Lemma 6. Suppose not; then, there is a stationary SPE f inducing an agreement $z \in \Delta^2$ at some finite $T (\leq 2)$. By standard perfection arguments, the responder j at T must be indifferent between accepting and rejecting z . Thus, by stationarity of f , $z_j = (1 - \delta)u_j(a) + \delta\pi_j^{T+1}$, where $a \in A^*$ is the disagreement outcome at T if z is rejected and π_j^{T+1} is the continuation payoff at $T + 1$.

Next, consider j deviating at T by not paying ρ . His continuation payoff at T is then $(1 - \delta)u_j(a') + \delta\pi_j^{T+1}$ for some $a' \in A^*$. But, since j 's equilibrium continuation payoff at T is $z_j - (1 - \delta)\rho = (1 - \delta)[u_j(a) - \rho] + \delta\pi_j^{T+1}$, and $\rho > b$, it follows that non-participation at T is profitable for j , which is a contradiction. \parallel

Proof of Proposition 6. By Lemma 1, if part (i) of the claim is true, part (ii) must also be true. To show (i), suppose otherwise. Then, there exists a SPEM M^* in the costly bargaining game inducing an agreement $z = (z_1, z_2) \in \Delta^2$ at some finite $T \geq 2$.

Suppose that i is the proposer at T . By Lemmas 1 and 2, j 's continuation payoff (given M_i^*) after rejecting any offer in any period at which j is the responder is at most $\delta^2(z_j - \rho) \geq 0$. This implies that, if M_j^* receives an offer $z' = (z_i + \epsilon, z_j - \epsilon) \in \Delta^2$ such that $(1 - \delta^2)z_j + \delta^2\rho > \epsilon > 0$, by subgame-perfection, it must always accept.

Now, consider another machine $M'_i = \{M'_{ip}, M'_{ir}\}$ which is identical to M_i^* except

that $\lambda'_{ip}(q_i^T, (\mathcal{I}, \mathcal{I})) = z'$. Since M_j^* always accepts z' , M'_i (given M_j^*) generates the same outcome path up to $t - 1$ as M_i^* and then agreement on z' at $t = \min\{\tau \leq T \mid q_i^\tau = q_i^T\}$. This improves i 's payoff, and hence, we have a contradiction. \parallel

References

- [1] Abreu, D., and A. Rubinstein, The Structure of Nash Equilibria in Repeated Games with Finite Automata, *Econometrica*, 56 (1988), 1259-1282.
- [2] Anderlini, L. and L. Felli, Costly Bargaining and Renegotiation, *Econometrica*, 69 (2001), 377-411.
- [3] Binmore, K., M. Piccione, and L. Samuelson, Evolutionary Stability in Alternating-Offers Bargaining Games, *Journal of Economic Theory*, 80 (1998), 257-291.
- [4] Bloise, G., Strategic Complexity and Equilibrium in Repeated Games, unpublished doctoral dissertation, University of Cambridge, 1998.
- [5] Busch, L-A., and Q. Wen, Perfect Equilibria in a Negotiation Model, *Econometrica*, 63 (1995), 545-565.
- [6] Chatterjee, K., and H. Sabourian, N-Person Bargaining and Strategic Complexity, *mimeo*, University of Cambridge, 1999.
- [7] Chatterjee, K., and H. Sabourian, Multiperson Bargaining and Strategic Complexity, *Econometrica*, 68 (2000), 1491-1509.
- [8] Coase, R. H., The Problem of Social Cost, *Journal of Law and Economics*, 3 (1960), 1-44.
- [9] Fernandez, R., and J. Glazer, Striking for a Bargain Between Two Completely Informed Agents, *American Economic Review*, 81 (1991), 240-252.
- [10] Gale, D., and H. Sabourian, Complexity and Competition, *Econometrica*, 73 (2005), 739-770.
- [11] Haller, H., and S. Holden, A Letter to the Editor on Wage Bargaining, *Journal of Economic Theory*, 52 (1990), 232-236.
- [12] Kalai, E., and W. Stanford, Finite Rationality and Interpersonal Complexity in Repeated Games, *Econometrica*, 56 (1988), 397-410.

- [13] Lee, J., and H. Sabourian, Complexity and Efficiency in Repeated Games with Negotiation, *Cambridge Working Papers in Economics*, 0419 (2004).
- [14] Lee, J., and H. Sabourian, Efficiency in Negotiation: Complexity and Costly Bargaining, *Birkbeck Working Papers in Economics and Finance*, 0505 (2005).
- [15] Piccione, M., Finite Automata Equilibria with Discounting, *Journal of Economic Theory*, 56 (1992), 180-193.
- [16] Piccione, M., and A. Rubinstein, Finite Automata Play a Repeated Extensive Game, *Journal of Economic Theory*, 61 (1993), 160-168.
- [17] Rubinstein, A. (1982): "Perfect Equilibrium in a Bargaining Model," *Econometrica*, 50, 97-109.
- [18] Rubinstein, A., Finite Automata Play the Repeated Prisoner's Dilemma, *Journal of Economic Theory*, 39 (1986), 83-96.
- [19] Sabourian, H., Bargaining and Markets: Complexity and the Competitive Outcome, *Journal of Economic Theory*, 116 (2003), 189-228.