Efficiency in Negotiation: Complexity and Costly Bargaining*

Jihong Lee
Birkbeck College, London†

Hamid Sabourian
King's College, Cambridge‡

Abstract

Even with complete information, two-person bargaining can generate a large number of equilibria, involving disagreements and inefficiencies, in (i) negotiation games where disagreement payoffs are endogenously determined (Busch and Wen [4]) and (ii) costly bargaining games where there are transaction/participation costs (Anderlini and Felli [2]). We show that when the players have (at the margin) a preference for less complex strategies only *efficient* equilibria survive in negotiation games (with sufficiently patient players) while, in sharp contrast, it is only the most *in*efficient outcome involving perpetual disagreement that survives in costly bargaining games. We also find that introducing small transaction costs to negotiation games dramatically alters the selection result: perpetual disagreement becomes the only feasible equilibrium outcome. Thus, in both alternating-offers bargaining games and repeated games with exit options (via bargaining and contracts), complexity considerations establish that the Coase Theorem is valid if and only if there are no transaction/participation costs.

JEL Classification: C72, C78
Keywords: Bargaining, Repeated Game, Coase Theorem, Transaction Cost, Complexity, Bounded Rationality, Automaton

1 Introduction: The Negotiation and Costly Bargaining Games

Under well-defined property rights, rational economic agents are expected to bargain and fully exploit any mutual gains from trade. This "Coase Theorem" (Coase [7]) provides an important benchmark for economists to think about the potential sources of inefficient outcomes of negotiation. Chief among many explanations for its failure is informational asymmetry, as documented by numerous papers in the literature on bargaining with incomplete information.[1]

Even with complete information, the Coase Theorem can be invalid. First, inefficiencies can be sustained as equilibria in "negotiation games" with complete information (Fernandez and Glazer [8], Haller and Holden [9] and Busch and Wen [4]).

The following defines the negotiation game of Busch and Wen (henceforth BW). There are two players indexed by $i = 1, 2$. In the alternating-offers protocol, each player in turn (player 1 in odd periods and player 2 in even periods) proposes a partition of a *periodic* surplus whose value

*The full version of this paper is available from http://www.econ.bbk.ac.uk/wp/ef/efwp.htm.

†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]For a survey of this literature, see Ausubel, Crampton and Deneckere [3].

is normalized to one.[2] 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.[3]

These models can be thought of either as bargaining games in which disagreement payoffs at each stage of the bargaining are determined endogenously in some game or as repeated games in which at each period there is an exit option via bargaining and contractual agreement.[4]

The negotiation game generally admits a large number (continuum) of subgame-perfect equilibria, as summarized by BW in a result that has a same flavor as the Folk theorem in repeated games. Some of these equilibria involve delay in agreement (even perpetual disagreement) and inefficiency. Thus, the Coase Theorem fails in the sense that it is no longer guaranteed.

Second, similar inefficiencies also arise in complete information bargaining models with small transaction costs (Anderlini and Felli [2], henceforth AF). In these models the players have to incur some (small) cost in order to participate in bargaining; such participation (or transaction) costs can induce sub-optimality for similar reasons as in the hold-up literature.

Two players engage in Rubinstein bargaining over a surplus of one (which accrues just once, not periodically) with player 1 making offers at odd dates and player 2 at even dates. There is no disagreement game to be played after a rejection. However, at the beginning of each period t, each player i has to pay a participation cost, denoted by $\rho \in [0, \frac{1}{2}]$, to enter the bargaining.

There are several ways to think about how this decision is made. The players can pay the cost either simultaneously or sequentially. As it turns out, the exact extensive form is immaterial to AF’s equilibrium characterization. Let us assume that the cost is paid sequentially. In particular, we assume that at each date 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.

Once both players have sunk the participation cost, the proposer makes an offer, followed by the other player’s response. The game ends in case of an acceptance; otherwise it moves onto the next period. If at least one player does not pay ρ, the game moves directly onto $t + 1$ without bargaining. We shall refer to this game as the ”costly bargaining game”.

A common feature in these explanations for inefficient bargaining is the multiplicity of equilibria. In particular, the games feature inefficient outcomes among a large number of equilibria some of which are also efficient. Such multiplicity of equilibria makes it difficult to draw firm conclusions concerning the Coase Theorem. In this paper, we take the two aforementioned approaches and attempt to select amongst the set of equilibria, thereby addressing the Coase Theorem head-on, by explicitly considering the complexity of implementing a strategy. In contrast to the literature on repeated games played by automata, we find that complexity considerations result in a very sharp set of (opposing) predictions in the above two approaches with regard to efficiency and the Coase Theorem.

Furthermore, in order to obtain a broader understanding of the role of transaction costs, we look at the “costly negotiation game’. Extending AF, we assume that in order for the players to

---

[2]Let $\triangle^2 \equiv \{x = (x_1, x_2) \mid \sum_i x_i = 1\}$ be a partition of the unit (periodic) surplus.

[3]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 in the disagreement game. We shall denote the set of action profiles in G by $A = A_1 \times A_2$ with its element indexed by a.

[4]A special case of this game is the standoff between a union and a firm considered by Fernandez and Glazer [8] (and Haller and Holden [9]). During a contract renewal process, a union and a firm renegotiate over the distribution of a periodic revenue, but a disagreement puts them in a strategic situation. After rejecting the firm’s wage offer or having their own offer rejected by the firm, the union can forego the status quo wage for one period and strike before a counter-offer is made next period. (The firm is inactive in the disagreement game.)

bargain in each period of the negotiation game (but not to play the disagreement game) both have to (sequentially) pay a participation cost; if at least one player foregoes the payment, they proceed directly to the disagreement game.

2 Strategy, Machines, Complexity and Equilibrium

There are many different ways of defining the complexity of a strategy in dynamic games. In the literature on repeated games played by automata the number of states of the machine is often used as a measure of complexity (Rubinstein [15], Abreu and Rubinstein [1], Piccione [13] and Piccione and Rubinstein [14]). This is because the set of states of the machine can be regarded as a partition of possible histories. In particular, Kalai and Stanford [10] show that the counting-states measure of complexity, or *state complexity*, is equivalent to looking at the number of *continuation strategies* that the strategy induces at different histories of the game. We extend this notion of strategic complexity to the negotiation game and the costly bargaining game, and facilitate the analysis by considering equivalent "machine games".

The alternating-offers bargaining imposes an asymmetric structure on the extensive forms considered which are stationary only every two periods (henceforth we shall refer to every two periods as a "stage"). To account for such structural asymmetry of the games, we adopt machine specifications that formally distinguishes between the different *roles* played by each player in a given stage. A player can be either proposer or responder. For most of the paper, a machine is assumed to consist of two "sub-machines", each playing a role (of proposer or responder) with distinct states, output and transition functions. Transition occurs at the end of each period, from a state belonging to one sub-machine to a state belonging to the other sub-machine as roles are reversed.[5]

Here are some notations:

$k = p, r$ indexes a player's "role" (proposer or responder)
D is the set of "partial histories" within a period
$D_{ik} = \{d \in D \mid \text{it is } i\text{'s turn to play in role } k \text{ after } d \text{ in the period}\}$
E is the set of all possible outcomes in a period
C_i is the set of actions available to player i
$C_{ik}(d)$ is the set of actions available to player i given k and $d \in D_{ik}$.

In order to represent a strategy in the above games, we employ a particular machine specification that consists of two "sub-machines":

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

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

[5]We show that the result of Kalai and Stanford [10], on the equivalence of the number of states and the number of continuation strategies that the implemented strategy induces, also holds here for our machine specifications.

Let Φ_i denote the set of player i's machines in the machine game. We 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 roles are reversed each period. We also assume that each sub-machine has to have at least one state.

Notice that we do not impose any restriction on the set of machines/strategies; each sub-machine may have any arbitrary (possibly infinite) number of states. This is in contrast to Abreu and Rubinstein [1] and others who consider only finite automata. Assuming that machines can only 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. Then, we adopt the following definition of strategic complexity:

Definition 2 (State complexity) *A machine M_i' is more complex than another machine M_i if $\|M_i'\| > \|M_i\|$.*

The following defines a *minimal* machine.

Definition 3 *A machine is minimal if and only if each of its sub-machines has exactly one state.*

A minimal machine implements the same actions in every period regardless of the history of the preceding periods, provided that the partial history within the current period (given a role) is the same.

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. We choose an equilibrium notion in which complexity enters a player's preferences *after* the payoffs and with a (non-negative) fixed cost c.[6]

To facilitate this concept, we first define the notion of ϵ-best response.

Definition 4 *For any $\epsilon \geq 0$, a machine M_i is a ϵ-best response to M_{-i} if, $\forall M_i'$,*

$$\pi_i(M_i, M_{-i}) + \epsilon \geq \pi_i(M_i', M_{-i}) \ .$$

If a machine is a 0-best response, then it is a best response in the conventional sense.

We then define a Nash equilibrium of the machine game with complexity cost c.

Definition 5 *A machine profile $M^* = (M_1^*, M_2^*)$ constitutes a Nash equilibrium of the machine game with complexity cost $c \geq 0$ (NEMc) if $\forall i$*

(i) M_i^* is a best response to M_{-i}^*;
(ii) $\exists$ no $M_i' \in \Phi_i$ such that M_i' is a c-best response to M_{-i}^* and $\|M_i^*\| > \|M_i'\|$.

By definition, the set of NEMc is a subset of the set of Nash equilibria in the negotiation game. The case of zero complexity cost $c = 0$ is closest to the standard equilibrium and corresponds to the case in which complexity enters players' preferences *lexicographically.* Any NEMc with a positive complexity cost $c > 0$ must also be a NEMc with $c = 0$. The magnitude of c therefore can be interpreted as a measure of how much the players care for less complex strategies, or indeed the players' *bounded rationality.*

[6] Sabourian [16] employs this equilibrium notion.

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

Definition 6 *A machine profile $M^* = (M_1^*, M_2^*)$ constitutes a subgame-perfect equilibrium of the machine game with complexity cost $c \geq 0$ (SPEMc) if M^* is both a NEMc and a subgame-perfect equilibrium (SPE) of the negotiation game.*

Payoffs In order to define payoffs in the negotiation game, lt us fix some more notational conventions. Let $M = (M_1, M_2)$ be a machine profile. Then, if M is the chosen machine profile, $T(M)$ refers to the end of the negotiation game; $z(M) \in \triangle^2$ is the agreement offer if $T(M) < \infty$; $a^t(M) \in A$ is the disagreement game outcome in period $t < T(M)$; and $q_i^t(M)$ is the state of player i's machine appearing in period $t \leq T(M)$ (the state of the active sub-machine in period t).

We then denote by $\pi_i^t(M)$ player i's (discounted) average *continuation payoff* at period $t \leq T(M)$ when the machine profile M is chosen.[7]

The following describes i's payoffs in the negotiation game:

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

In order to define payoffs for the costly bargaining game, we first denote by $\rho_i(M)$ the sum of participation costs that player i pays along the entire outcome path induced by machine profile M, discounted at the appropriate rate. If f induces an agreement $z = (z_1, z_2) \in \triangle^2$ in a finite period T, then the payoff to i is given by

$$\pi_i(M) = \delta^{T-1} z_i - \rho_i(M)$$

while if the strategies induce perpetual disagreement then player i's payoff is given by

$$\pi_i(M) = -\rho_i(M)\ .$$

3 Results

We show that when the players have a preference for less complex strategies at the margin, only *efficient* equilibria survive in negotiation games while it is only the most *in*efficient equilibrium involving perpetual disagreement that survives in bargaining games with (arbitrarily small) transaction costs. We also combine the two approaches by introducing transaction costs to negotiation games. We find that the selection result for negotiation games with complexity is dramatically altered in the presence of (arbitrarily small) transaction costs: perpetual disagreement becomes the only feasible equilibrium outcome.[8].

[7] We shall use the abbreviation $\pi_i(M) = \pi_i^1(M)$.

[8] To save space we omit formal proofs and refer the reader to our working paper (Lee and Sabourian [11])

3.1 The Negotiation Game

Some Preliminary Results We begin by laying out some Lemmas that will pave way for the main results below. We first state an obvious, yet very important, implication of the complexity requirement. Suppose that there exists a state in some player's equilibrium (NEMc) machine that never appears on the equilibrium path. Unless the machine is minimal, however, this cannot be possible because this state can be "dropped" by the player to reduce complexity cost without affecting the outcome and payoff, thereby contradicting the NEMc assumption. This argument leads to the following Lemma.

Lemma 1 *Assume that* $M^* = (M_1^*, M_2^*)$ *is a NEMc, where* $M_i^* = \{M_{ip}^*, M_{ir}^*\}$ *and* $M_{ik}^* = (Q_{ik}^*, q_{ik}^{1*}, \lambda_{ik}^*, \mu_{ik}^*)$ *for any* $i = 1, 2$ *and* $k = p, r$.[9] *Then, we have the following:*
(i) if $T(M^*) \geq 2$*, then* $\forall i$*,* $\forall k$ *and* $\forall q_i \in Q_{ik}^*$ $\exists$ *a period* t *such that* $q_i^t(M^*) = q_i$*;*
(ii) if $T(M^*) \leq 2$*, then* $|Q_{ik}| = 1$ $\forall i$ *and* $\forall k$.[10]

Next note that, since any strategy can be implemented by a machine, it follows, by definition, that any NEMc profile $M^* = (M_1^*, M_2^*)$ corresponds to a Nash equilibrium of the underlying negotiation game; thus

$$\pi_i(M_i^*, M_j^*) = \max_{f_i \in F_i} \pi_i(f_i, M_j^*) \quad \forall i, j$$

where, with some abuse of notation, $\pi_i(f_i, M_j^*)$ refers to i's payoff in the game where i and j play according to f_i and M_j^* respectively.

More generally, the equilibrium machines must be best responses (in terms of payoffs) *along* the equilibrium path of the negotiation game.

Lemma 2 *Assume that* $M^* = (M_1^*, M_2^*)$ *is a NEMc. Then,* $\forall i, j$ *and* $\forall \tau \leq T(M^*)$ *we have*

$$\pi_i^\tau(M^*) = \max_{f_i \in F_i^\tau} \pi_i(f_i, M_j^*(q_j^\tau))$$

where $q_j^\tau \equiv q_j^\tau(M^*)$*,* $M_j^*(q_j^\tau) \in \Phi_j^\tau$ *is the machine that is identical to* M_j^* *except that it starts with the sub-machine which operates in period* τ *with initial state* q_j^τ*, and again with some abuse of notation,* $\pi_i(f_i, M_j^*(q_j^\tau))$ *refers to i's payoff in the negotiation game that starts with role distribution given 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.*

It then follows that if a state belonging to a player's equilibrium machine appears twice on the outcome path then the continuation payoff of the other player must be identical at both periods.

Lemma 3 *Assume that* $M^* = (M_1^*, M_2^*)$ *is a NEMc with* $c \geq 0$*. Then,* $\forall i, j$ *and* $\forall t, t' \leq T(M^*)$ *we have the following:*

$$\text{if} \;\; q_j^t(M^*) = q_j^{t'}(M^*), \;\text{then}\; \pi_i^t(M^*) = \pi_i^{t'}(M^*)\ .$$

The proofs of these lemmas are simple adaptations of the earlier results to the 2SM machine specification;

[9] This will henceforth define the equilibrium machines in our claims.

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

Agreement We now show that, if an agreement occurs at some finite period as a NEMc outcome, then it must occur within the very first stage (two periods) of the negotiation game, and thus, the associated equilibrium machines (strategies) must be minimal (stationary).

To do so, we first establish the following critical Lemma: if a NEMc induces an agreement in a finite period then every state of the equilibrium machines occurs only once on the equilibrium path. The proof uses the arguments behind Lemmas 3 and 4 in AR which deliver their "tracking states" result. In AR, 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 but when there is an agreement in a finite period cycles clearly cannot happen.

Lemma 4 *Assume that* $M^* = (M_1^*, M_2^*)$ *is a NEMc with* $T(M^*) < \infty$*. Then,* $q_i^t(M^*) \neq q_i^{t'}(M^*)$ $\forall t, t' \leq T(M^*)$ *and* $\forall i$*.*

We are now ready to present our first major result. Any NEMc outcome that reaches an agreement must do so in the very first stage of the negotiation game and hence the associated strategies must be stationary. The intuition is as follows. We know from Lemma 4 that the state of each player's machine occurring in the last period is distinct. This implies that, if the last period occurs beyond the first stage of the game, one of the players must be able to drop his last period's state without affecting the outcome of the game. In the case where the final offer on the equilibrium path was proposed before by the same player, he could reduce complexity cost simply by using one state to make the offer twice. On the other hand, if the final offer occurs only once then the responder in the last period could reduce complexity cost by replacing the last period's state with any other state in his (sub-)machine and then 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 machine in a given state can be a function of the proposal.[11]

Proposition 1 *Assume that* $M^* = (M_1^*, M_2^*)$ *is a NEMc with* $T(M^*) < \infty$*. Then, (i)* $T(M^*) \leq 2$*, and (ii)* M_1^* *and* M_2^* *are minimal and hence* M^* *is stationary.*

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

Corollary 1 *For any* $\epsilon \in (0, 1)$*,* $\exists\, \bar{\delta} < 1$ *such that for any* $\delta \in (\bar{\delta}, 1)$*, any NEMc profile* M^* *of the negotiation game with discount factor* δ *that involves an agreement must be such that* $\sum_i \pi_i(M^*) > 1 - \epsilon$*.*

Stationary Subgame-Perfect Equilibria We begin the SPEMc characterization of the negotiation game by considering *stationary* SPE. Since our notion of a stationary strategy allows for actions conditional on partial history within a period, a stationary SPE here does not precisely correspond to BW's characterization (see their Proposition 1 and Corollary 1).

Of course, it must be 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.

[11] Chatterjee and Sabourian [5][6] use similar intuition to derive a similar result in a multi-person 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 [6], is based on a different notion of complexity from one considered here.

But, a stationary SPE here is not necessarily efficient. Delay in agreement (either over one period or indefinite) and inefficiency can be sustained in equilibrium because a player who makes a deviating offer can be credibly punished in the disagreement game of the same period if the disagreement game has multiple Nash equilibria.

Before providing a characterization for the set of stationary subgame-perfect equilibria, denote the set of Nash equilibria of G by A^* and let

$$b \equiv \max_i \sup_{a,a' \in A^*} [u_i(a) - u_i(a')].^{12} \tag{1}$$

Proposition 2 *(i) The negotiation game has a stationary SPE if and only if A^* is non-empty.*
(ii)If f is a stationary SPE profile of the negotiation game, then

$$\sum_i \pi_i(f) \geq 1 - (1-\delta)b \ .$$

It immediately follows from Proposition 2 that for a sufficiently large discount factor every stationary SPE must be (almost) efficient.

Corollary 2 *For any $\epsilon \in (0,1)$, $\exists\ \bar{\delta} < 1$ such that, for any $\delta \in (\bar{\delta}, 1)$, if f is a stationary SPE of the negotiation game with discount factor δ then $\sum_i \pi_i(f) > 1 - \epsilon$.*

We can also deduce that in some cases a stationary SPE is efficient independently of the discount factor.

Corollary 3 *For any δ, every stationary SPE of the negotiation game is efficient if either (i) G has a unique Nash equilibrium or (ii) $\forall a \in A^*$ we have $\sum_i u_i(a) < 1 - b$.*

SPEMc and Perpetual Disagreement First, let $\Omega^\delta(c)$ denote the set of SPEMc machine profiles in the negotiation game with discount factor δ and complexity cost c.

Now we show that, given a discount factor arbitrarily close to one, any SPEMc 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.

The proof of this claim (and also several other claims below) requires an interaction between complexity and perfection arguments. The key complexity argument is that every state of each player's equilibrium machine must appear on the equilibrium path (Lemma 1). This implies the following. Suppose that a player deviates from a SPEMc of the negotiation game by making a different offer in some period. What can the other player obtain if he rejects this offer? Since the state of each player's (sub-)machine is fixed for each period (not at each decision node), the ensuing disagreement game of the period may see an outcome that never happens on the original equilibrium path; but then, Lemma 1 implies that the subsequent transition must take the players to some point along the original path for next period. Thus, any punishment for a player who deviates from the proposed equilibrium must itself occur on the equilibrium path (except for the play of the disagreement game immediately after the deviating offer), and as a consequence, the set of equilibrium outcomes is severely restricted.

Informally, we consider the period in which a player gets his maximum continuation payoff in the proposer role. Bargaining can then be used by the other player in the *preceding* period to break

[12] Notice that $b \in [0,1]$.

up the on-going disagreement if there is any (continuation) inefficiency from then on. In such cases, there exists a Pareto-improving deviation offer because the responder in that period, who will be proposing next, cannot obtain more from punishing the deviant than what he is already getting from the original outcome as of next period. We need the discount factor to be sufficiently large so as to eliminate the importance of the current period in which the deviation can be followed immediately by an off-the-equilibrium play of the disagreement game.

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

Proposition 3 does not however rule out the possibility that we observe inefficiency (in terms of payoffs) early on in the negotiation game.[13] Given any $\epsilon > 0$ and δ sufficiently close to one, we can write the total equilibrium payoff from the negotiation game as

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

where M^* is the equilibrium profile ($T(M^*) = \infty$) and τ is the first period in which continuation becomes (almost) efficient. The limit of the right-hand side as $\epsilon \to 0$ and $\delta \to 1$ is not necessarily the efficient level (because τ may depend on δ).[14]

Main Results for SPEMc Now, putting together Proposition 1, Proposition 3 and Corollary 1, we state Theorem 1: under sufficiently patient players, (i) every SPEMc (in fact any NEMc) inducing an agreement must do so in the very first stage of the negotiation game and hence be stationary and almost efficient, (ii) every SPEMc inducing perpetual disagreement must be at least almost efficient in the long run, and (iii) if the structure of the disagreement game is such that there exists no efficient action profile, the players cannot disagree forever; every SPEMc must then induce an agreement in the first stage and hence be stationary and almost efficient.

Theorem 1 *1. For any δ and any $c \geq 0$, if $M^* \in \Omega^\delta(c)$ is such that $T(M^*) < \infty$ then $T(M^*) \leq 2$ and M^* is stationary.*

2. For any $\epsilon \in (0,1)$, $\exists\ \bar{\delta} < 1$ such that, for any $\delta \in (\bar{\delta}, 1)$ and any $c \geq 0$, we have

(a) if $M^* \in \Omega^\delta(c)$ is such that $T(M^*) < \infty$ then $\sum_i \pi_i(M^*) > 1 - \epsilon$;

(b) if $M^* \in \Omega^\delta(c)$ is such that $T(M^*) = \infty$ then $\exists\ \tau < \infty$ such that $\sum_i \pi_i^\tau(M^*) > 1 - \epsilon$;

(c) if $\sum_i u_i(a) < 1\ \forall a \in A$ then every $M^* \in \Omega^\delta(c)$ is such that $T(M^*) \leq 2$, and hence is stationary, and $\sum_i \pi_i(M^*) > 1 - \epsilon$.

In fact, when complexity cost c is strictly positive, we obtain a sharper efficiency result: for a sufficiently large δ, every SPEMc of the negotiation game must be stationary and hence almost efficient however small that complexity cost is.

[13] To be precise, neither does it rule out the possibility that there will be inefficient disagreement game outcomes even after τ. It is just that the continuation game from then on is almost efficient.

[14] If we restrict each player's machine to use only a finite number of states, then any machine profile must generate *cycles*. But this is not enough to guarantee that Proposition 3 implies ex ante efficiency in the limit as δ goes to one. For this, we need for instance to additionally assume that the size of a machine is uniformly bounded (for any δ) so that the first cycle cannot last beyond a fixed period.

Theorem 2 *For any $c > 0$, we have the following: for any $\epsilon \in (0, 1)$, $\exists\ \bar{\delta} < 1$ such that, for any $\delta \in (\bar{\delta}, 1)$, every $M^* \in \Omega^{\delta}(c)$ is stationary and such that $\sum_i \pi_i(M^*) > 1 - \epsilon$.*

We know from Corollary 2 that (almost) efficiency follows from stationarity. The intuition for stationarity in above Theorem is as follows (for sufficiently patient players). If equilibrium machines are not stationary then by Proposition 1 there cannot be any agreement. Next, consider the (first) period, say τ_η, at which a player, say 2, gets his maximum continuation payoff in the proposer role. Bargaining can then be used by the other player in the preceding period to break up the on-going disagreement and save the state at τ_η. In particular, since 2 gets his maximum continuation payoff at date τ_η, there is a deviation offer by player 1 at date $\tau_\eta - 1$ that is acceptable by player 2 and involves an arbitrarily small loss for 1. Since such deviation induces an agreement at $\tau_\eta - 1$ and thus saves the state at τ_η and complexity cost is positive, it follows that 1 is better-off deviating.[15]

We can further relate the set of SPEMc in the negotiation game to the structure of the disagreement game G. For instance, Corollary 3 reports some cases where a stationary SPE of the negotiation game is efficient independently of the discount factor. In those cases, we have a stronger set of efficiency results.

Corollary 4 *Suppose either G has a unique Nash equilibrium[16] or $\sum_i u_i(a) < 1 - b\ \forall a \in A^*$. Then, every $M^* \in \Omega^{\delta}(c)$ is efficient (i.e. $\sum_i \pi_i(M^*) = 1$) if $T(M^*) < \infty$. Moreover, $\exists\ \bar{\delta} < 1$ such that for any $\delta \in (\bar{\delta}, 1)$ every $M^* \in \Omega^{\delta}(c)$ is efficient if either $\sum_i u_i(a) < 1\ \forall a \in A$ or $c > 0$.*

3.2 The Costly Bargaining Game

Let us now consider how complexity considerations affect the set of equilibria in the costly bargaining game. We begin by making the following observation.

Lemma 5 *Fix any $\rho > 0$, and assume that $M^* = (M_1^*, M_2^*)$ is a NEMc in the costly bargaining game such that $T(M^*) = \infty$. Then, (i) neither player pays ρ in any period and (ii) M_i^* is minimal for all i.*

Note that Lemmas 1-3 above also hold here in the costly bargaining game. To save space we shall omit their statements and proofs.

Next it is obvious that the unique stationary SPE, in which neither player pays ρ in any period and bargaining never takes place, is also a SPEMc.[17] We now show that this is indeed the unique SPEMc if $\rho > 0$. To demonstrate this, we first show that, for any SPEMc, if there is an agreement it must be at period 1. The basic arguments here are similar to those for Proposition 3 in the previous section. Because of complexity considerations any deviation can be punished only by what happens on the equilibrium path. This implies in this bargaining setup that, if there is a delay in agreement and a deviating offer is made in the last period, rejection leads to a strictly lower continuation payoff for the responder (because of discounting and/or positive participation cost). It then follows from subgame-perfection that there must exist a deviating offer for the last period's proposer which the responder will accept and improves his payoff.

[15] As in the case of Proposition 3, we need δ to be sufficiently large so as to eliminate the importance of the current period in which the deviation can be followed immediately by an off-the-equilibrium play of the disagreement game.

[16] This is true, for example, in the union-firm negotiation models of Fernandez and Glazer [8] and Haller and Holden [9] in which only one player (the union) acts in the disagreement game and there is a single strictly dominant action (no strike).

[17] This is because stationary strategies can be implemented by minimal machines.

Proposition 4 *Fix any $\rho > 0$, and assume that M^* is a SPEMc in the costly bargaining game such that $T(M^*) < \infty$. Then, $T(M^*) = 1$.*

Combining Lemma 5 and Proposition 4, we now state our next Theorem. Complexity together with any positive participation cost selects a unique equilibrium outcome in two-person bargaining, which is extremely *in*efficient.

Theorem 3 *For any $\rho > 0$, every SPEMc profile M^* in the costly bargaining game is such that (i) $T(M^*) = \infty$, (ii) neither player pays ρ in any period, and (iii) M_i^* is minimal for all i.*

3.3 The Costly Negotiation Game

Let us now investigate the impact of transaction costs on our selection results in the negotiation game. We consider the following "costly negotiation game".

Extending the ideas of AF, we assume that in order for the players to enter bargaining (but *not* the disagreement game) in each period both must sequentially pay a cost $\rho \in [0, \frac{1}{2}]$ at the beginning of each period. In odd (even) periods, player 1 (2) first decides whether or not to pay ρ. If he makes the payment, player 2 (1) then decides whether or not to pay ρ. Bargaining in that period occurs if and only if both players sink the cost; otherwise the game moves directly to the disagreement game before moving onto the next period.

With appropriate modifications to the notations, all previous definitions and notations in the previous section on histories, strategies, complexity and machines also carry to the costly negotiation game. Also, let ρ_i^t be the (discounted) sum of participation costs that player i incurs between periods t and T under profile M. Now we can define player i's (discounted) *average* continuation payoff at period t when M is chosen as

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

We begin the NEMc/SPEMc characterization here by noting that by exactly the same reasoning as in the previous section the properties of NEMc profiles specified in Lemmas 1-4 above also hold for the costly negotiation game. To save space we shall omit their statements and proofs.

It turns out that the introduction of the transaction cost dramatically alters our selection results. We first establish, as in Proposition 1 above, that if an agreement takes place in some finite period as a NEMc outcome of the costly negotiation game it must be in the first stage. The proof builds on from some of the arguments behind Proposition 1.

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

We are now ready to state our next Theorem. Let, as before, A^* be the set of Nash equilibria in G and define $b \equiv \max_i \sup_{a,a' \in A^*}[u_i(a) - u_i(a')]$.

Theorem 4 *Every SPEMc profile M^* in the costly negotiation game is such that:*

1. *if $\rho > b$ then $T(M^*) = \infty$ and neither player pays ρ in any period;*
2. *if $0 < \rho \leq b$ then either $T(M^*) \leq 2$ or $T(M^*) = \infty$.*

We can then immediately claim the following Corollaries.

Corollary 5 *Suppose that we have* $\sum_i u_i(a) < 1\ \forall a \in A$. *Then, for any* $\rho > b$, *every SPEMc of the costly negotiation game is inefficient.*

Corollary 6 *Suppose that* G *has a unique Nash equilibrium. Then, for any* $\rho > 0$, *every SPEMc* M^* *of the costly negotiation game is such that (i)* $T(M^*) = \infty$, *(ii) neither player pays* ρ *in any period, and (iii) if* $\sum_i u_i(a) < 1\ \forall a \in A$ *then the outcome is inefficient.*

Notice the stark contrast between the above two corollaries and Theorem 1/Corollary 4 above for the case in which every disagreement game outcome is dominated by an agreement (i.e. $\sum_i u_i(a) < 1\ \forall a \in A$). In particular, when in addition G has a unique Nash equilibrium, the only feasible SPEMc outcome in the negotiation game with no participation cost is an agreement in the very first stage which is therefore efficient (in the limit); with the participation cost, we only have perpetual disagreement and *in*efficiency.

4 Conclusion

Our results suggest the following:

(i) The Coase Theorem continues to be valid in negotiation games with bargaining and endogenously determined disagreement payoffs. Since such negotiation models can also be thought of as repeated games with exit options (see below) we can alternatively interpret our results as showing that complexity and bargaining in tandem offer a sharp explanation for co-operation in repeated games;

(ii) Transaction costs are a critical ingredient of a *robust* explanation for perpetual disagreements and inefficient negotiation outcomes under complete information.

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] Ausubel, L. M., P. Crampton, and R. J. Deneckere, Bargaining with Incomplete Information, in R. J. Aumann and S. Hart (eds.), *Handbook of Game Theory*, Vol. 3, Elsevier, Amsterdam, 2002.

[4] Busch, L-A., and Q. Wen, Perfect Equilibria in a Negotiation Model, *Econometrica*, 63 (1995), 545-565.

[5] Chatterjee, K., and H. Sabourian, N-Person Bargaining and Strategic Complexity, *mimeo*, University of Cambridge, 1999.

[6] Chatterjee, K., and H. Sabourian, Multiperson Bargaining and Strategic Complexity, *Econometrica*, 68 (2000), 1491-1509.

[7] Coase, R. H., The Problem of Social Cost, *Journal of Law and Economics*, 3 (1960), 1-44.

[8] Fernandez, R., and J. Glazer, Striking for a Bargain Between Two Completely Informed Agents, *American Economic Review*, 81 (1991), 240-252.

[9] Haller, H., and S. Holden, A Letter to the Editor on Wage Bargaining, *Journal of Economic Theory*, 52 (1990), 232-236.

[10] Kalai, E., and W. Stanford, Finite Rationality and Interpersonal Complexity in Repeated Games, *Econometrica*, 56 (1988), 397-410.

[11] Lee, J., and H. Sabourian, Efficiency in Negotiation: Complexity and Costly Bargaining, *Birkbeck Working Papers in Economics and Finance*, 0505 (2005)

[12] Osborne, M., and A. Rubinstein, *A Course in Game Theory*, MIT Press, Cambridge, Massachusetts, 1994.

[13] Piccione, M., Finite Automata Equilibria with Discounting, *Journal of Economic Theory*, 56 (1992), 180-193.

[14] Piccione, M., and A. Rubinstein, Finite Automata Play a Repeated Extensive Game, *Journal of Economic Theory*, 61 (1993), 160-168.

[15] Rubinstein, A., Finite Automata Play the Repeated Prisoner's Dilemma, *Journal of Economic Theory*, 39 (1986), 83-96.

[16] Sabourian, H., Bargaining and Markets: Complexity and the Competitive Outcome, *Journal of Economic Theory*, 116 (2003), 189-228.