ISSN 1745-8587

Birkbeck UNIVERSITY OF LONDON

School of Economics, Mathematics and Statistics

Birkbeck Working Papers in Economics & Finance

BWPEF 0505

Efficiency in Negotiation: Complexity and Costly Bargaining

Jihong Lee
Hamid Sabourian

January 2005

▪ Birkbeck, University of London ▪ Malet Street ▪ London ▪ WC1E 7HX ▪

Efficiency in Negotiation: Complexity and Costly Bargaining

Jihong Lee
Birkbeck College, London*

Hamid Sabourian
King's College, Cambridge†

January 2005

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 [6]) 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

*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

Under well-defined property rights, rational economic agents are expected to bargain and fully exploit any mutual gains from trade. This "Coase Theorem" (Coase [9]) 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 [10], Haller and Holden [13] and Busch and Wen [6]). 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. Inefficiencies in these games take the form of delay in agreement and even perpetual disagreement. Second, similar inefficiencies also arise in complete information bargaining models with small transaction costs (Anderlini and Felli [2]). 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.[2]

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.

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. Thus, our results suggest the following:

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

[2] Other explanations for possible inefficiencies in complete information models can be found in Perry and Reny [18], Sákovics [23] and Fershtman and Seidmann[11].

(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.

We first examine the "negotiation game" of Busch and Wen [6] (or simply BW). In each period, two players bargain - in Rubinstein's alternating-offers protocol - over the distribution of a fixed and commonly known periodic surplus. If an offer is accepted, the game ends and each player gets his share of the surplus according to the agreement in every period thereafter. After any rejection, but before the game moves to the next period, the players engage in a normal form game to determine their payoffs for the period. The Pareto frontier of the disagreement game is contained in the bargaining frontier.

A special case of this game is the standoff between a union and a firm considered by Fernandez and Glazer [10] (and Haller and Holden [13]). 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.)

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.

As we mentioned before, the negotiation game and its equilibria can be interpreted from two alternative viewpoints. Naturally we can think of the game as a standard alternating-offers bargaining game with endogenous disagreement payoffs and hence evaluate its equilibria in a Coasian context.[3] The alternative viewpoint focuses on the repeated game aspect of the game. Real world repeated interactions are often accompanied by negotiations which can lead to mutual agreement. While equilibria in standard repeated games are usually given the interpretation of implicit, self-enforcing agreements, the situations depicted by the negotiation game are associated with the possibility of explicit contracts that can bind the players to a particular set of outcomes. For example, we observe firms engaged in a repeated horizontal or vertical relationship negotiating

[3]The issue of endogenous disagreement payoffs in a bargaining situation goes back at least to Nash [16] who considers the problem in a co-operative framework.

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.

The Folk theorem gives economic theorists little hope of making any predictions in repeated interactions. However, as the aforementioned examples suggest, it seems that negotiation is often a salient feature of real world repeated interactions, presumably to enforce co-operation and efficient outcomes. Can bargaining be used to isolate equilibria in repeated games? Unfortunately, the contributions of BW and others demonstrate that Folk theorem type results are likely to persist even when the players are endowed with an opportunity at the beginning of each period to settle on an efficient outcome once and for all.

We next consider the "costly bargaining game" of Anderlini and Felli [2] (or simply AF). Two players bargain in the Rubinstein alternating-offers protocol to split a fixed surplus (which accrues once, not periodically as in the negotiation game), but at the beginning of each period both players have a choice of whether or not to pay a participation cost. (It is immaterial to the results whether the cost is paid simultaneously or sequentially). Bargaining in that period takes place only if both players pay their participation cost. If at least one player decides not to pay, there will be no bargaining and the game moves to the next period.

AF characterize the set of subgame-perfect equilibria in this game. In particular, they show that when the participation cost is low, the game admits a wide variety of equilibria. Immediate agreement, agreement with an arbitrarily long delay, and no bargaining/no agreement (with players not paying the participation cost) at all dates are all possible equilibrium outcomes.

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 [21], Abreu and Rubinstein [1], Piccione [19] and Piccione and Rubinstein [20]). This is because the set of states of the machine can be regarded as a partition of possible histories. In particular, Kalai and Stanford [14] 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. [4,5]

The concept of Nash equilibrium is then refined to incorporate the players' preference for less complex strategies. In our choice of equilibrium, complexity enters a player's preferences, together with the payoffs in the underlying game, either lexicographically or as a positive fixed cost c. The larger this cost is, the more is required of complexity. We refer to a Nash equilibrium (of the machine game) with fixed complexity cost c by NEMc and adopt the convention of using $c = 0$ (and thus NEM0) to refer to the lexicographic case in which players first choose a best response and then choose the least complex strategy within the set of best responses. We also invoke the notion of subgame-perfection and consider the set of NEMc that are subgame-perfect, referred to as SPEMc.

The selection results are as follows. We first show that complexity considerations select only efficient equilibria in the negotiation game. Independently of the complexity cost and discount factor, if an agreement occurs in some finite period as the outcome of some NEMc then it must occur within the very first stage of the game, and moreover, the equilibrium strategies are stationary (history-independent). Thus, in this case any NEMc outcome is efficient in the limit as the discount factor goes to one. We then show that, with sufficiently patient players, every SPEMc in the negotiation game that induces perpetual disagreement is at least long-run almost efficient; that is, the players must reach a finite period in which the continuation game then on is almost efficient.[6] For the case of positive complexity cost, we establish a stronger result; every SPEMc of the negotiation game is stationary and hence almost efficient, with sufficiently patient players.[7]

Second, we show that, in sharp contrast to the previous results, complexity selects the most inefficient outcome in the costly bargaining game. Independently of the complexity cost, discount factor and participation cost, every SPEMc in the costly bargaining game

[4]We also explore an alternative machine specification that employs more frequent transitions and hence account for finer partitions of histories and continuation strategies. This machine consists of four sub-machines; while keeping the role distinction, transition occurs twice in each period at the end of bargaining and at the end of the disagreement game.

[5]We show that the result of Kalai and Stanford [14], 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.

[6]This implies that we can draw a yet stronger set of conclusions under certain disagreement game structures (given sufficiently high discounting). For example, if every disagreement game outcome is dominated by an agreement, it is not possible to have a SPEMc involving perpetual disagreement. Thus, every equilibrium outcome in this case must reach an agreement in the first stage of the negotiation game, and hence is stationary and almost efficient.

[7]With the four sub-machine specification mentioned in footnote 4, we derive the same set of SPEMc results independently of the discount factor.

is such that the players never pay the participation cost and therefore agreement is never reached.

Finally, 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 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.

The striking observation here is that the introduction of the participation cost dramatically alters the selection result for the negotiation game. We show how just a small positive participation cost induces perpetual disagreement in every SPEMc of the costly negotiation game. This result thus strengthens our conclusion that transaction costs are indeed a critical ingredient of a robust account of the failure of the Coase Theorem under complete information.[8]

In terms of new ideas and techniques about complexity in games, the paper's main contributions lie in the following. First, we successfully combine some of the ideas developed in looking at complexity in repeated games (Abreu and Rubinstein [1]) and alternating-offers bargaining models (Chatterjee and Sabourian [7][8]) to a setting that involves bargaining, repeated games and/or transaction costs.[9] Second, we introduce a new machine specification that involves different sub-machines to play different roles. Such a specification is critical for our results and may also be useful in other dynamic games.[10] Moreover, with this specification the counting state measure of complexity is equivalent to Kalai and Stanford's [14] notion of complexity. Third, we show how perfection and complexity arguments intricately interact to deliver the results on perpetual disagreement.

In terms of application, the results are novel and unique in that they contain very sharp and clear predictions on the issues of agreement and efficiency. For example, taken from the repeated game perspective, the efficiency results in the negotiation game take the study of complexity in repeated games a step further from the existing literature in which complexity has yielded only a limited selective power. While many inefficient

[8] Again the results here are sharper if $c > 0$ or if the four sub-machine specification is adopted.

[9] We like to point out two differences between our approach here and the above cited literature. First, we do not assume that machines have a finite number of states. Second, our complexity definition is different from that used in the recent literature on complexity and bargaining; for example, Chatterjee and Sabourian [8] look at complexity of the response rule within a period. (See also Sabourian [22] and Gale and Sabourian [12] who use a similar notion of complexity to justify competitive outcomes in market games with matching and bargaining.)

[10] In the previous literature, this issue did not arise because the analysis was restricted either to repeated games (and therefore there was a natural specification of the machines) or to pure bargaining problems. Here, since the negotiation game is a mixture of repeated game and bargaining, having a machine specification with different sub-machines playing different roles turns out to play an important role.

equilibria survive complexity refinement in standard repeated games (Abreu and Rubinsten [1]),[11] we demonstrate that complexity and bargaining in tandem can provide a sharp explanation for why co-operation and efficiency can be expected to arise in repeated interactions. On the other hand, our results also show that introducing (small) transaction costs to the standard bargaining or negotiation game can swing the selection results sharply towards perpetual disagreement and inefficiency.

The paper is organized as follows. We begin Section 2 by describing the negotiation game and BW's main results. We then introduce the notion of complexity in terms of strategies and machines, describe the machine game and finally present the main analysis and efficiency results for the negotiation game. In Section 3 we run the analogous analysis for the bargaining and negotiation games with transaction costs. In sharp contrast to the results in Section 2 with no transaction costs, we learn here that complexity selects perpetual disagreement and inefficient outcomes in the presence of transaction costs in both games. We conclude in Section 4. Appendix contains some relegated proofs.

2 Complexity in the Negotiation Game

2.1 The Basics

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 $t = 1, 2, \ldots$ and adopt the convention that player 1 makes offers in odd periods and player 2 makes offers in even periods. Let $\triangle^2 \equiv \{x = (x_1, x_2) \mid \sum_i x_i = 1\}$ be a partition of the unit periodic surplus. A period then refers to a single offer $x \in \triangle^2$ 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 $u_i(\cdot) : A_1 \times A_2 \to R$ is his payoff function in the disagreement game. We shall denote the set of action profiles

[11] Abreu and Rubinstein [1] show that complexity refinement narrows the set of equilibrium payoffs in 2×2 repeated games. But, in other repeated games complexity by itself does not have any bite at all. (See Bloise [5] who shows robust examples of two-player repeated games in which the set of Nash equilibria with complexity costs coincides with the set of individually rational payoffs.)

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, the bargaining offers the players an opportunity to settle on an efficient outcome once and for all.

Two types of outcome paths are possible in the negotiation game; one in which an agreement occurs in a finite time and one in which disagreement continues perpetually. Let T denote the end of the game and $a^t \in A$ the disagreement game outcome (action profile) in period $t < T$. If $T = \infty$, we mean an outcome path in which agreement is never reached. Player i's (discounted) *average* payoff in this case is equal to

$$(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}u_i(a^t) \ .$$

If $T < \infty$, denote the agreed partition in T by $z = (z_1, z_2) \in \triangle^2$. Player i's payoff from such an outcome path amounts to

$$(1-\delta)\sum_{t=1}^{T-1}\delta^{t-1}u_i\left(a^t\right) + \delta^{T-1}z_i \ .$$

The negotiation game is stationary only every two periods (beginning with an odd one) or "stage". In specifying the players' strategies (and later machines), we shall formally distinguish between the different *roles* played by each player in each stage game. He can be either the proposer (p) or the responder (r) in a given period. We shall index a player's role by k. The role distinction provides a natural framework to capture the structural asymmetry that the alternating-offers bargaining imposes on the repeated (disagreement) game.

In order to define a strategy, we first need to introduce some further notations. We shall use the following notational convention. 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 $k \neq l$.

We shall denote by e a history of outcomes in a period, and this belongs to the set

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

where the superscript i represents the identity of the proposer in the period. Let e^t be the outcome of the period t.

We also need notations 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, we define a "partial history" (information within a period) d as an element in the following set

$$D = \{\emptyset, (x^i), (x^i, N)\}_{x^i \in \triangle^2, i=1,2} \ .$$

For example, the null set $\emptyset$ here refers to the beginning of a period at which the proposer has to make an offer; (x^i, N) represents a partial history of an offer x^i by player i followed by the other player's rejection. Also, let us define

$$D_{ik} \equiv \{d \in D \mid \text{it is } i\text{'s turn to play in role } k \text{ after } d \text{ in the period}\} .$$

Thus, we have

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

We denote the set of actions available to player i by

$$C_i \equiv \triangle^2 \cup Y \cup N \cup A_i .$$

Let $C_{ik}(d)$ be the set of actions available to player i given his role k and a corresponding partial history $d \in D_{ik}$. Thus, we have

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

Let

$$H^t = \underbrace{E \times \cdots \times E}_{t \text{ times}}$$

be the set of all possible histories of outcomes over t periods in the negotiation game, excluding those that have resulted in an agreement. The initial history is empty (trivial) and denoted by $H^1 = \emptyset$. Let $H^\infty \equiv \cup_{t=1}^\infty H^t$ denote the set of all possible finite period histories.

For the analysis, we shall divide H^∞ into two smaller subsets according to the different roles that the players play in each stage. Let H_{ik}^t be the set of all possible histories over t periods after which player i's role is k. Notice that $H_{ik}^t = H_{jl}^t$. Also, let $H_{ik}^\infty = \cup_{t=1}^\infty H_{ik}^t$. Thus, $H^\infty = H_{ip}^\infty \cup H_{ir}^\infty$ $(i = 1, 2)$.

A strategy for player i is then a function

$$f_i : (H_{ip}^\infty \times D_{ip}) \cup (H_{ir}^\infty \times D_{ir}) \to C_i$$

such that for any $(h, d) \in H_{ik}^\infty \times D_{ik}$ we have $f_i(h, d) \in C_{ik}(d)$. The set of all strategies for player i is denoted by F_i. Also, we shall 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, $F_i^t = F_i$.

We can define a *stationary* (or history-independent) strategy in the following way.

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

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 the period. For instance, if a strategy f_i is stationary, then it must be such that

$$f_i(h, x, N) = f_i(h', x, N) \ \ \forall h, h' \in H^{\infty} \ \forall x \in \triangle^2$$

but we may have

$$f_i(h, x, N) \neq f_i(h, x', N) \ \text{ for } x \neq x' \ .$$

Notice also that a stationary strategy profile always induces the same outcome in each stage of the game.

In the spirit of the Folk theorem, BW characterize the set of all subgame-perfect equilibrium (SPE) payoffs of the negotiation game. BW, to this end, compute the lower bound of each player's SPE payoff in the negotiation game with discount factor δ. Define

$$w_j = \max_{a \in A} \left\{ u_j(a) - \left[\max_{a'_i \in A_i} u_i(a'_i, a_j) - u_i(a) \right] \right\}$$

which BW assume to be well-defined. Note also that $w_i \leq 1$ given the assumption that $u_i(a) \leq 1 \ \forall a \in A$, and that $w_i \geq 0$ if G has at least one Nash equilibrium.

Then, BW show that the infimum of player i's SPE payoffs in the negotiation game beginning with his offer (given δ) is not less than $\underline{v}_i(\delta) = \frac{1-w_j}{1+\delta}$ while the infimum of the other player's SPE payoffs in the same game is not less than $\underline{v}_j(\delta) = \frac{\delta(1-w_i)}{1+\delta}$. They also show that, with sufficiently patient players, there exists a SPE of the negotiation game (beginning with i's offer) in which the players obtain these lower bounds.

Next define the limit of these infima as δ goes to unity by

$$\underline{v}_i = \frac{1 - w_j}{2} \ \text{ and } \ \underline{v}_j = \frac{1 - w_i}{2} \ .$$

We can now formally state the BW's main Theorem.

BW Theorem *For any payoff vector* (v_1, v_2) *of the negotiation game such that* $v_1 > \underline{v}_1$ *and* $v_2 > \underline{v}_2$*,* $\exists \ \bar{\delta} \in (0,1)$ *such that* $\forall \delta \in (\bar{\delta}, 1)$*,* (v_1, v_2) *is a SPE payoff vector of the negotiation game with discount factor* δ*.*

The forces of bargaining thus restrict the set of feasible equilibrium payoffs in the negotiation game compared to the set of individually rational payoffs in the disagreement (repeated) game. But, if $\underline{v}_1 + \underline{v}_2 < 1$, the negotiation game has many inefficient subgame-perfect equilibria much in the way the Folk theorem characterizes the repeated game (even when the disagreement game payoffs are always uniformly small relative to agreement). The negotiation game has a unique (efficient) SPE payoff if $\underline{v}_1 + \underline{v}_2 = 1$ or $w_1 = w_2 = 0$ which implies that any Nash equilibrium payoff vector of the disagreement game has to coincide with its minmax point.

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 [14]).

In a repeated game, it is natural to take the measure over all its possible subgames. In the negotiation game, each stage game consists of an extensive form game and this means that several different definitions are possible.

In the main analysis, we shall consider the set of all continuation strategies at the beginning of each period of the negotiation game. Formally, let $f_i|h$ be the continuation strategy at history $h \in H^\infty$ induced by $f_i \in F_i$. Thus,

$$f_i|h(h', d) = f_i(h, h', d) \text{ for any } (h, h', d) \in H_{ik}^\infty \times D_{ik} \text{ for any } k \ .$$

Also, let us define the set of all such continuation strategies by $F_i(f_i) = \{f_i|h : h \in H^\infty\}$. Then the cardinality of this set provides a measure of strategic complexity. Let us call it $comp(f_i)$.

The set of continuation strategies can also be divided into smaller sets according to the role specification. Define $F_{ik}(f_i) = \{f_i|h : h \in H_{ik}^\infty\}$ such that we have $F_i(f_i) = \cup_k F_{ik}(f_i)$. Complexity can then be equivalently measured by $comp(f_i) = \sum_k |F_{ik}(f_i)|$.

We can also measure complexity over finer partitions of histories and corresponding continuation strategies. We shall later show that the exact definition of complexity is going to play some role in shaping the precise details of the results.

In dynamic games any strategy can be *implemented* by an automaton or a "machine" (we shall clarify this statement below in our negotiation game context). Moreover, Kalai and Stanford [14] (or simply KS) show that in repeated games 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, 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 that this equivalence between the two representations of strategic complexity also holds in the negotiation game. Our approach to complexity will then be facilitated in machine terms as this will provide a more economical platform to present the analysis of complexity. 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 interchangeably use with the negotiation game.

Here the extensive form of the stage game allows for many different machine specifications to equivalently represent a strategy. (The same is also the case in other sequential dynamic games; see Piccione and Rubinstein [20], Chatterjee and Sabourian [7][8] and Sabourian [22]). The fact that the stage game is also asymmetric across its two periods - a player switches his role in the bargaining process - adds to this issue of multiple possible machine specifications.[12]

In this paper, we will mainly employ a particular machine specification that consists of two "sub-machines":

Definition 2 *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 $k, l = p, r$*

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

Let Φ_i denote the set of player i's machines in the machine game. We also let Φ^t_i 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, $\Phi^t_i = \Phi_i$.

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.[13,14]

[12] There is no loss of generality in adopting any of the possible specifications in terms of implementing strategies. However, as we shall see below, the counting-the-number-of-states (of a machine) measure of complexity corresponds to different notions of complexity for different machine specifications.

[13] We could also define a distinct terminal state for each sub-machine. This is immaterial. We are assuming that if an offer is accepted by the responder, M_i enters the terminal state of the relevant sub-machine and shuts off.

[14] Note 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. Nevertheless, we endow both sub-machines with an initial state for expositional ease.

Let us now formally state what we mean by a machine implementing a strategy in the negotiation game. Consider a machine $M_i = \{M_{ip}, M_{ir}\} \in \Phi_i$ where, for $k = p, r$, $M_{ik} = (Q_{ik}, q^1_{ik}, \lambda_{ik}, \mu_{ik})$. For every $k = p, r$ and for any $h \in H^\infty_{ik}$, denote the state at history h by $q_i(h) \in Q_{ik}$. Formally if $h = (e^1, \ldots, e^{t-1})$ then $q_i(h) = q^t_i$ where, for any $0 < \tau \leq t$, q^τ_i is defined inductively by

$$q^1_i = \begin{cases} q^1_{ik} & \text{if } i \text{ is in role } k \text{ initially at } t = 1 \\ q^1_{il} & \text{if } i \text{ is in role } l \text{ initially at } t = 1 \end{cases}$$

$$q^\tau_i \equiv \begin{cases} \mu_{il}(q^{\tau-1}_i, e^{\tau-1}) & \text{if } i \text{ is in role } k \text{ at } \tau \\ \mu_{ik}(q^{\tau-1}_i, e^{\tau-1}) & \text{if } i \text{ is in role } l \text{ at } \tau \end{cases}, \qquad \forall \tau > 0$$

Definition 3 *M_i implements f_i if $\forall k$, $\forall h \in H^\infty_{ik}$ and $d \in D_{ik}$ we have*

$$\lambda_{ik}(q_i(h), d) = f_i(h, d)$$

where $q_i(h)$ is defined inductively as above.

Clearly, any strategy f_i can be implemented by a machine. For example, consider a machine $M_i = \{M_{ip}, M_{ir}\}$ (where, for $k = p, r$, $M_{ik} = (Q_{ik}, q^1_{ik}, \lambda_{ik}, \mu_{ik})$) which is such that $\forall k$, $\forall h \in H^\infty_{ik}$, $\forall d \in D_{ik}$ and $\forall e \in E$

$$Q_{ik} = H^\infty_{ik},\ \lambda_{ik}(h, d) = f_i(h, d),\ \mu_{ik}(h, e) = (h, e),$$

and $q^1_{ik'} = \emptyset$ where k' is i's role in period $t = 1$. Evidently, this machine implements f_i.

The above establishes that machines and strategies are equivalent in our set-up. Notice also 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. Next, as in KS (Theorem 1), we show that for any strategy f_i, $comp(f_i)$ 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.[15]

[15]The exact specification of a machine is important in qualifying this statement. Since in defining $comp(f_i)$ 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. It is also important that each sub-machine uses its own distinct set of states.

Proposition 1 *Fix any $f_i \in F_i$. Let $\mathcal{M}_i(f_i) \subseteq \Phi_i$ be the set of machines that implement f_i. Also, let $\bar{M}_i = \{\bar{M}_{ip}, \bar{M}_{ir}\}$ be a smallest machine that implements f_i; that is*

$$\|\bar{M}_i\| \in \{M_i \in \mathcal{M}_i(f_i) \mid \|M_i\| \leq \|M_i'\| \ \forall M_i' \in \mathcal{M}_i(f_i)\} \ .$$

Then, we have $|F_{ik}(f_i)| = \|\bar{M}_{ik}\|$ for any $k = p, r$ and thus $\|\bar{M}_i\| = comp(f_i)$.

Proof. See Appendix. ‖

Given this result, we now formally define the notion of complexity in terms of machines, as in Rubinstein [21] and Abreu and Rubinstein [1].[16]

Definition 4 (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 5 *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. Thus, it corresponds to a stationary strategy as in Definition 1. We shall henceforth refer to a minimal machine (profile) interchangeably as a stationary machine (profile).

To wrap up the description of the machine game, let 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).

Similarly, we 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. Thus,

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

[16] We also draw attention to the work of Binmore, Piccione, and Samuelson [4] who propose another notion of complexity similar to state complexity considered in this paper and others. According to their "collapsing state condition", an automaton M^1 is less complex than another automaton M^2 if the same implementation can be obtained by consolidating a collection of states belonging to M^2 into a single state in M^1. It will not be difficult to see that our results will also hold under this notion of complexity.

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

For ease of exposition, the argument in M will sometimes be dropped when we refer to one of these variables that depends on the particular machine profile. For example, π_i^t will refer to $\pi_i^t(M)$. Unless otherwise stated, the abbreviated variable will refer to the machine profile in the *claim*.

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.[17]

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

Definition 6 *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 7 *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*.

Abreu and Rubinstein [1] (or simply AR) propose a general way of describing a player's preference ordering over machine profiles that is increasing in his payoff of the game and decreasing in the complexity of his machine. A Nash equilibrium can then be written in terms of machines that are most preferred against each other. In contrast, our equilibrium concept directly finds a subset of Nash equilibria of the underlying game that fits our complexity cost criterion (at the margin). There is, however, an analytical parallel between our choice of solution concept and that of AR because the latter must also be a Nash equilibrium of the underlying negotiation game (see Lee and Sabourian [15]). Our complexity cost criterion can be thought of as an alternative way to embed the trade-off between payoff and complexity that underlies AR's preference ordering.

[17] Sabourian [22] employs this equilibrium notion.

NEMc strategy profiles are not necessarily credible however. We could introduce credibility, as in Chatterjee and Sabourian [7][8], by introducing trembles into the model and considering 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. The trembles will ensure that strategies are optimal (allowing for complexity) at every information set that occurs with a positive probability.

A more direct, and simpler, way of introducing credibility would be to consider NEMc strategy profiles that are subgame-perfect equilibria of the negotiation game without complexity cost.

Definition 8 *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.*

Given Proposition 1, we can equivalently define these notions of equilibrium (NEMc and SPEMc) 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. Also, unless otherwise stated explicitly, any reported result and its proof below are independent of (i) the complexity cost ($c \geq 0$), including the lexicographic case, and (ii) the discount factor δ.

2.3 Analysis: Complexity and Efficiency

2.3.1 NEMc Results

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 $k = p, r$.*[18] *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 $\forall k$.[19]

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

[19] 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.

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:*

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

Lemmas 1-3 are familiar arguments rooted in the existing literature (see Abreu and Rubinstein [1] or Osborne and Rubinstein [17]). The proofs of these lemmas are simple adaptations of the earlier results to the 2SM machine specification; hence to save space we shall omit them and refer the reader to our working paper (Lee and Sabourian [15]).

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 $\forall i$.*

***Proof*.** Suppose not; then there exists a NEMc profile M^* with $T(M^*) < \infty$ such that $q_i^t = q_i^{t'}$ for some i and for two distinct dates $t, t' \leq T$. Now, we show that there has to be a cycle, which contradicts that $T(M^*) < \infty$.

Define τ_i as the minimal t such that q_i^t is repeated. 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 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 $M_{jk}' = (Q_{jk}', q_{jk}^{1\prime}, \lambda_{jk}', \mu_{jk}')$ for $k = p, r$, which is identical to M_j^* except that (i) $q_j^{\tau_i}$ is dropped (thus $Q_{jl}' = Q_{jk}^* \backslash q_j^{\tau_i}$) 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 Lemma 3, $\pi_j(M_i^*, M_j') = \pi_j(M^*)$. But then, since $q_j^{\tau_i}$ is dropped, $\|M_j^*\| > \|M_j'\|$, and this contradicts NEMc.[20]

Step 2: $\tau_i = \tau_j$ where τ_j is the minimal t such that q_j^t is repeated.

By Step 1 $\tau_j \leq \tau_i$. 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' be the minimal $t > \tau_j$ such that $q_j^t = q_j^{\tau_j}$.

Suppose not; suppose $\tau_i' > \tau_j'$. Let $\tau = \tau_i = \tau_j$. There are two cases to consider.

<u>*Case A*</u>: $q_j^{\tau_i'} \neq q_j^t$ $\forall t < \tau_i'$.

Now, consider another machine $M_j' = \{M_{jp}', M_{jr}'\}$ which is identical to M_j^* except that (i) $q_j^{\tau_i'}$ is dropped (thus $Q_{jl}' = Q_{jl}^* \backslash q_j^{\tau_i'}$) 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:

[20] Notice that this result turns on the assumption that each sub-machine uses a distinct set of states. If the sub-machines shared the states, we could not simply "drop" $q_j^{\tau_i}$ since it could be used for the other sub-machine (playing a different role) somewhere else.

$$
\begin{aligned}
\pi_j^\tau(M_i^*, M_j') &= (1-\delta)\left[\sum_{t=\tau}^{\tau_i'-1} \delta^{t-\tau} u_j(a^t) + \delta^{\tau_i'-\tau} \sum_{t=\tau}^{\tau_i'-1} \delta^{t-\tau} u_j(a^t) + \dots\right] \\
&= \frac{1-\delta}{1-\delta^{\tau_i'-\tau}} \sum_{t=\tau}^{\tau_i'-1} \delta^{t-\tau} u_j(a^t)\ .
\end{aligned}
$$

But this equals to $\pi_j^\tau(M_i^*, M_j^*)$ because

$$
\begin{aligned}
\pi_j^\tau(M_i^*, M_j^*) &= (1-\delta)\sum_{t=\tau}^{\tau_i'-1} \delta^{t-\tau} u_j(a^t) + \delta^{\tau_i'-\tau} \pi_j^{\tau_i'}(M_i^*, M_j^*) \\
&= (1-\delta)\sum_{t=\tau}^{\tau_i'-1} \delta^{t-\tau} u_j(a^t) + \delta^{\tau_i'-\tau} \pi_j^{\tau}(M_i^*, M_j^*) \\
&= \frac{1-\delta}{1-\delta^{\tau_i'-\tau}} \sum_{t=\tau}^{\tau_i'-1} \delta^{t-\tau} u_j(a^t)\ ,
\end{aligned}
$$

where the second equality follows from Lemma 3.

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, $\|M_j^*\| > \|M_j'\|$, and this contradicts NEMc.

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 (thus $Q_{il}' = Q_{il}^* \backslash q_i^\tau$) and (ii) the transition function is such that $\mu_{il'}'(q_i^{\tau-1}, e^{\tau-1}) = q_i^{\tau_j'}$ and $\mu_{il'}'(q_i^{\tau_i'-1}, e^{\tau_i'-1}) = q_i^s$, where i play 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 get

$$
\pi_i^\tau(M_i', M_j^*) = (1-\delta)\sum_{t=\tau_j'}^{\tau_i'-1} \delta^{t-\tau_j'} u_i(a^t) + \frac{(1-\delta)\delta^{\tau_i'-\tau_j'}}{1-\delta^{\tau_i'-s}} \sum_{t=s}^{\tau_i'-1} \delta^{t-s} u_i(a^t)\ .
$$

But, since $q_j^\tau = q_j^{\tau'_j}$ and $q_j^{\tau'_i} = q_j^s$, Lemma 3 implies

$$
\begin{aligned}
\pi_i^\tau(M^*) &= \pi_i^{\tau'_j}(M^*) \\
&= (1-\delta)\sum_{t=\tau'_j}^{\tau'_i-1} \delta^{t-\tau'_j} u_i(a^t) + \delta^{\tau'_i-\tau'_j}\pi_i^{\tau'_i}(M^*) \\
&= (1-\delta)\sum_{t=\tau'_j}^{\tau'_i-1} \delta^{t-\tau'_j} u_i(a^t) + \delta^{\tau'_i-\tau'_j}\pi_i^{s}(M^*) \\
&= (1-\delta)\sum_{t=\tau'_j}^{\tau'_i-1} \delta^{t-\tau'_j} u_i(a^t) + \frac{(1-\delta)\delta^{\tau'_i-\tau'_j}}{1-\delta^{\tau'_i-s}}\sum_{t=s}^{\tau'_i-1} \delta^{t-s} u_i(a^t)
\end{aligned}
$$

where the last equality follows from

$$\pi_i^{\tau'_i} = \pi_i^s = (1-\delta)\sum_{t=s}^{\tau'_i-1} \delta^{t-s} u_i(a^t) + \delta^{\tau'_i-s}\pi_i^{\tau'_i} .$$

Therefore, $\pi_i(M'_i, M^*_j) = \pi_i(M^*)$. But since $\|M^*_i\| > \|M'_i\|$ we have a contradiction.

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 $T < \infty$. $\|$

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.[21]

[21] Chatterjee and Sabourian [7][8] use similar intuition to derive a similar result in a multi-person

Proposition 2 *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.*

Proof. If part (i) of the claim is true, part (ii) must be true because of Lemma 1. Let us consider part (i).

Suppose not. Then an agreement $z \in \triangle^2$ occurs at some $T \in (2, \infty)$. We know from Lemma 4 that q_1^T and q_2^T are both distinct. Now suppose that player i is the proposer at T and consider two possible cases.

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) the transition function is such that $\mu_{ir}'(q_i^{T-1}, e^{T-1}) = q_i^\tau$. (Note that $q_i^\tau \neq q_i^T$ since we have $T > 2$ and q_i^T is distinct by Lemma 4.)

Since $\lambda_{ip}'(q_i^\tau, \emptyset) = z$ and q_i^T appears for the first time at T on the original equilibrium path, M_i' (given M_j^*) generates an identical outcome path and payoff as M_i^*. But, since q_i^T is dropped, $\|M_i^*\| > \|M_i'\|$. This contradicts NEMc.

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) the transition function is such that $\mu_{jp}'(q_j^{T-1}, e^{T-1}) = q_j \neq q_j^T$ for some arbitrary but fixed $q_j \in Q_{jr}'$ (such q_j exists since we have $T > 2$ and q_j^T is distinct by Lemma 4) and (iii) the output function is such that $\lambda_{jr}'(q_j, z) = Y$.

Since the offer z does not appear anywhere before T on the original equilibrium path when i proposes, playing 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 NEMc. $\|$

It immediately follows from Proposition 2 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* $\sum_i \pi_i(M^*) > 1 - \epsilon$*.*

2.3.2 SPEMc Results

Stationary Subgame-Perfect Equilibria We begin the SPEMc characterization of the negotiation game by considering *stationary* SPE. Since our notion of a stationary strategy (Definition 1) allows for actions conditional on partial history within a period,

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 [8], is based on a different notion of complexity from one considered here.

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.

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.[22]

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

$$b \equiv \max_i \sup_{a,a' \in A^*} [u_i(a) - u_i(a')].^{23} \tag{3}$$

Proposition 3 *(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

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

[22]For instance, suppose that G has three Nash equilibria $a = (a_1, a_2)$, $a^1 = (a_1^1, a_2^1)$ and $a^2 = (a_1^2, a_2^2)$ such that the following two conditions hold

$$1 > \sum_i u_i(a) \geq 1 - (1-\delta) \max_i (u_i(a^i) - u_i(a)) \tag{1}$$

$$u_i(a^i) > u_i(a) \ \ \forall i \ . \tag{2}$$

Then, there exists a stationary SPE in which the players disagree indefinitely and play a in every period after rejection. We can easily check that given (1) and (2) the following stationary strategy profile $f = (f_1, f_2)$ constitutes a SPE. For each i, f_i is such that $\forall h \in H_{ip}^{\infty}$

$$f_i(h, \emptyset) = x^i = (x_1^i, x_2^i) \in \triangle^2 \text{ such that } x_j^i < u_j(a)$$
$$f_i(h, x, N) = \begin{cases} a_i & \text{if } x = x^i \\ a_i^j & \text{if } x \neq x^i \end{cases}$$

and $\forall h \in H_{ir}^{\infty}$

$$f_i(h, x) = Y \ \text{ if and only if } x_i > (1-\delta)u_i(a^i) + \delta u_i(a)$$
$$f_i(h, x, N) = \begin{cases} a_i & \text{if } x = x^j \\ a_i^i & \text{if } x \neq x^j \ . \end{cases}$$

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

Proof. See Appendix. $\|$

It immediately follows from Proposition 3 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 $\sum_i u_i(a) < 1 - b$.*

Proof. See Appendix. $\|$

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 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 4 *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 $\sum_i \pi_i^\tau(M^*) > 1 - \epsilon$.*

Proof. Fix any $\epsilon \in (0,1)$. Define

$$\beta = \max\left\{1, \max_i \sup_{a,a' \in A} [u_i(a) - u_i(a')]\right\} \tag{4}$$

which is bounded since $u(\cdot)$ is. Define also

$$\bar{\delta} = 1 - \frac{\epsilon}{\beta}\ .$$

Given these, consider any $\delta \in (\bar{\delta}, 1)$ (thus $\epsilon > \beta(1-\delta)$) and any $M^* = (M_1^*, M_2^*) \in \Omega^\delta(c)$ with $T(M^*) = \infty$. Define η, t_{ik} and τ_η such that

$$0 < \eta < \epsilon - \beta(1-\delta), \tag{5}$$

$$t_{ik} = \{t |\ i \text{ plays role } k\}, \tag{6}$$

$$\tau_\eta = \min\{t \in t_{2p} \mid \pi_2^t + \eta > \pi_2^{t'}\ \forall t' \in t_{2p}\} (< \infty)\ . \tag{7}$$

Now, given M_1^*, consider player 2's continuation payoff after rejecting any offer in any period belonging to t_{2r}. Notice that since

- every state of M_1^* appears on the equilibrium path of M^* (Lemma 1)
- $\pi_2^t = \max_{f_2 \in F_2^t} (f_2, M_1^*(q_1^t))\ \forall t$ (Lemma 2),

player 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 $\pi_2^{\tau_\eta} + \eta > \pi_2^t$, $\forall t \in t_{2p}$.

The above implies that if 2's equilibrium machine M_2^* receives an offer

$$z' = (1 - \pi_{2r}^{\max}, \pi_{2r}^{\max}) \in \triangle^2,$$

where

$$\pi_{2r}^{\max} = (1-\delta) \sup_{a \in A} u_2(a) + \delta \left(\pi_2^{\tau_\eta} + \eta\right), \tag{8}$$

it must always accept because the profile M^* is subgame-perfect.

Now, consider player 1 using another machine $M_1' = \{M_{1p}', M_{1r}'\}$ which is identical to M_1^* except that $\lambda_{1p}'(q_1^{\tau_\eta - 1}, \emptyset) = z'$. Define

$$\tau = \min_t \{t | \ q_1^t = q_1^{\tau_\eta - 1}\} \ . \tag{9}$$

Since M_2^* always accepts the offer z' and M_1' differs from M_1^* only in offers conditional on state $q_1^{\tau_\eta - 1}$, it follows that (M_1', M_2^*) results in agreement z' in period τ.

We also know by Lemma 3 that $\pi_2^\tau = \pi_2^{\tau_\eta - 1}$. Thus, we have

$$\pi_2^\tau = (1-\delta) u_2(a^{\tau_\eta - 1}) + \delta \pi_2^{\tau_\eta} \ . \tag{10}$$

Since $\sup_{a \in A} u_2(a) - u_2(a^{\tau_\eta - 1}) \leq \beta$ (where β is given by (4)), we have, by (8) and (10),

$$\pi_{2r}^{\max} - \pi_2^\tau \leq (1-\delta)\beta + \delta\eta \ .$$

Using this, we can write

$$1 - \pi_{2r}^{\max} \geq 1 - (\pi_2^\tau + (1-\delta)\beta + \delta\eta) \ . \tag{11}$$

Since M^* is a SPEMc, it must be that $\pi_1^\tau \geq 1 - \pi_{2r}^{\max}$; otherwise the deviation to M_1' is profitable. This implies that (given $\delta < 1$)

$$\pi_1^\tau + \pi_2^\tau > 1 - ((1-\delta)\,\beta + \eta) \ .$$

But, since by (5) we have $\epsilon > (1-\delta)\,\beta + \eta$, it follows that at period $\tau < \infty$, $\sum_i \pi_i^\tau > 1 - \epsilon$ as in the claim. ‖

Proposition 4 does not however rule out the possibility that we observe inefficiency (in terms of payoffs) early on in the negotiation game.[24] 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).[25]

[24] 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.

[25] 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 4 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.

Main Results for SPEMc Now, putting together Proposition 2, Proposition 4 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.

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 $\sum_i \pi_i(M^*) > 1 - \epsilon$.*

Proof. See Appendix. ‖

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 2 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.[26]

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*[27] *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 $c > 0$.*

2.4 Discussion

Alternative Machine Specifications Since each stage game of the negotiation game has a sequential structure, we can have alternative machine specifications that employ more frequent transitions and hence account for finer partitions of histories and continuation strategies.

For example, below we have a definition of a machine, henceforth called the 4SM specification, that maintains the role distinction and employs distinct sub-machines to play the bargaining and the disagreement game within each period. Transition thus takes place twice within each period - once after the bargaining and once after the disagreement game.[28]

Definition 9 (Four sub-machine (4SM) specification) *A machine, $M_i = \{M_{ip}, \tilde{M}_{ip}, M_{ir}, \tilde{M}_{ir}\}$, consists of four sub-machines $M_{ik} = (Q_{ik}, q^1_{ik}, \lambda_{ik}, \mu_{ik})$ and $\tilde{M}_{ik} = (\tilde{Q}_{ik}, \tilde{q}^1_{ik}, \tilde{\lambda}_{ik}, \tilde{\mu}_{ik})$ for $k = p, r$. Each sub-machine consists of a set of states, an initial state, an output function and a transition function such that $\forall q_{ik} \in Q_{ik}$,*

[26] As in the case of Proposition 4, 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.

[27] This is true, for example, in the union-firm negotiation models of Fernandez and Glazer [10] and Haller and Holden [13] in which only one player (the union) acts in the disagreement game and there is a single strictly dominant action (no strike).

[28] We can also construct a machine in which transition occurs at each decision node of the stage game. Six sub-machines will then be required (some of which will in fact serve only to make transition and not output). There are several other possible specifications. But we conjecture that as long as we keep the role distinction for the bargaining part the results will remain.

$\forall \tilde{q}_{ik} \in \tilde{Q}_{ik}$, $\forall x^i, x^j \in \triangle^2$ *and* $\forall a \in A$,

$$\begin{array}{rl} \lambda_{ip}(q_{ip}, \emptyset) \in \triangle^2; & \lambda_{ir}(q_{ir}, x^j) \in \{Y, N\}; \\ \tilde{\lambda}_{ip}(\tilde{q}_{ip}, \emptyset) \in A_i; & \tilde{\lambda}_{ir}(\tilde{q}_{ir}, \emptyset) \in A_i; \\ \mu_{ip}(q_{ip}, x^i, N) \in \tilde{Q}_{ip}; & \mu_{ir}(q_{ir}, x^j, N) \in \tilde{Q}_{ir}; \\ \tilde{\mu}_{ip}(\tilde{q}_{ip}, a) \in Q_{ir}; & \tilde{\mu}_{ir}(\tilde{q}_{ir}, a) \in Q_{ip} \ . \end{array}$$

A minimal machine in the 4SM specification is again a machine whose sub-machines have only one state each, but it corresponds to an alternative notion of stationarity. In contrast to the 2SM case, a minimal machine here plays the disagreement game independently of partial history within a period. Therefore, it captures Markov-stationary behavior. Since, by Proposition 1 and Corollary 1 of BW, any Markov-stationary SPE induces an efficient outcome it follows that a minimal 4SM profile is efficient independently of the discount factor. Using this, we can derive for the 4SM specification (see Lee and Sabourian [15]) a set of NEMc/SPEMc results that contain much the same flavor as the corresponding results in the 2SM specification above, but are yet sharper: the discount factor no longer matters in the perpetual disagreement case.[29] More formally, let $\tilde{\Omega}^\delta(c)$ denote the set of SPEMc in 4SM given discount factor δ and complexity cost c. Then we can summarize the SPEMc results under the 4SM specification as follows.

Theorem 3 *Consider any δ and any $c \geq 0$. Then we have the following:*

1. *If $M^* \in \tilde{\Omega}^\delta(c)$ is such that $T(M^*) < \infty$, then $T(M^*) \leq 2$, M^* is Markov-stationary and hence efficient.*
2. *If $M^* \in \tilde{\Omega}^\delta(c)$ is such that $T(M^*) = \infty$, then for any $\epsilon > 0$ $\exists$ $\tau < \infty$ such that $\sum_i \pi_i^\tau(M^*) > 1 - \epsilon$.*
3. *If $\sum_i u_i(a) < 1$ $\forall a \in A$ then every $M^* \in \tilde{\Omega}^\delta(c)$ induces an immediate agreement, and hence is Markov-stationary and efficient.*
4. *If $c > 0$ then every $M^* \in \tilde{\Omega}^\delta(c)$ is Markov-stationary and hence is efficient.*

To save space we shall omit the proof of the above and refer the reader to Lee and Sabourian [15].

The No Discounting Case We also need to mention that our results are not affected by considering the negotiation game without discounting. In fact, by taking $\delta = 1$ (and using the limit of means criterion), a sharper result obtains in the perpetual disagreement case (Proposition 4). This case has to be (exactly) efficient in the long run.

[29] The definitions of NEMc/SPEMc are as before except that complexity (counting the number of states) of a 4SM machine now corresponds to the set of continuation strategies of the implemented strategy at the beginning of each period and at the beginning of the disagreement game in each period.

3 The Role of Transaction Costs

3.1 Costly Bargaining

AF investigate the impact of transaction costs in two-person negotiation by considering the following "costly bargaining game". Unless stated otherwise, the same set of notations will be used as before.

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.[30] 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.

Let $\mathcal{I}$ and $\mathcal{N}$ denote a player's decision to pay and not pay the participation cost respectively. (We sometimes refer to $\mathcal{I}$ simply as "participation".) In this game, a partial history within a period, d, belongs to the set

$$D = \{\emptyset, (\mathcal{I}), (\mathcal{I}, \mathcal{I}), (\mathcal{I}, \mathcal{I}, x^i)\}_{i=1,2,x^i \in \triangle^2}$$

where, for example, $(\mathcal{I})$ represents the partial history of payment of the participation cost by the player who proposes in the period and $(\mathcal{I}, \mathcal{I})$ the partial history of sequential payment of the cost by both players. Thus, we have

$$D_{ip} = \{\emptyset, (\mathcal{I}, \mathcal{I})\} \quad \text{and} \quad D_{ir} = \{(\mathcal{I}), (\mathcal{I}, \mathcal{I}, x^j)\}_{x^j \in \triangle^2} \ .$$

We similarly modify the definition of E, the set of outcomes in a period. Let e denote an element in this set.

The set of player i's actions is $C_i \equiv \mathcal{I} \cup \mathcal{N} \cup \triangle^2 \cup Y \cup N$. We also have

$$C_{ip}(d) = \begin{cases} \{\mathcal{I}, \mathcal{N}\} & \text{if } d = \emptyset \\ \triangle^2 & \text{if } d = (\mathcal{I}, \mathcal{I}) \end{cases}$$

$$C_{ir}(d) = \begin{cases} \{\mathcal{I}, \mathcal{N}\} & \text{if } d = (\mathcal{I}) \\ \{Y, N\} & \text{if } d = (\mathcal{I}, \mathcal{I}, x^j) \ . \end{cases}$$

[30]Every result and its proof in this subsection apply to the case of simultaneous payment.

Given these modifications, the definitions of histories and strategies remain the same as before. Let us define the payoffs. We denote by $\rho_i(f)$ the sum of participation costs that player i pays along the entire outcome path induced by strategy profile $f = (f_1, f_2)$, 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

$$\pi_i(f) = \delta_i^{T-1} z_i - \rho_i(f)$$

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

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

AF first show that the costly bargaining game admits a SPE in which neither player ever pays the participation cost and therefore disagreement persists forever. (Henceforth, we shall refer to this equilibrium as perpetual disagreement.) Using this SPE as a penal code, they demonstrate that there exists a continuum of equilibrium outcome paths. Furthermore, perpetual disagreement is the only stationary SPE.[31,32]

AF Theorem *In the costly bargaining game with positive participation cost ρ, we have the following:*[33]

1. *If ρ is sufficiently large ($\rho > \frac{\delta}{1+2\delta}$), perpetual disagreement is the unique SPE.*
2. *For other values of ρ ($\rho \leq \frac{\delta}{1+2\delta}$),*
   (a) *perpetual disagreement is a stationary SPE;*
   (b) *there exists a SPE that induces an immediate agreement at $x = (x_1, x_2) \in \triangle^2$ if and only if $x_1 \in [1 - \delta(1 - 2\rho), 1\text{-}\rho]$;*
   (c) *every $x \in \triangle^2$ that can be supported as an immediate agreement by a SPE can also be supported by a SPE involving an arbitrary length of delay.*
3. *Perpetual disagreement is the unique stationary SPE.*

Let us now consider how complexity considerations affect the set of equilibria in the costly bargaining game. We assume that machines have the 2SM specification as in the

[31] Otherwise, there is a stationary SPE with an agreement at some T such that the responder at T would be indifferent between accepting the final offer and waiting until the next period. But, in order to receive the offer, he would have paid the participation cost. It would then be strictly better for him to hold up the payment and wait, which gives a contradiction.

[32] AF 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.

[33] AF characterize the equilibria only for the case of simultaneous participation decision, but it is clear that the same qualitative statement can be made also for the sequential case.

previous section. Any strategy in this game can be implemented by a 2SM machine. All the other definitions and notations on complexity and machines remain the same and the result of Kalai and Stanford [14] (Proposition 1) equally extends.

We continue to investigate the set of NEMc and SPEMc. 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.*

Proof. (i) Since $T(M^*) = \infty$, this follows immediately from the observation that any player can always obtain a zero payoff by not participating in any period.

(ii) Suppose not; then M_i^* is not minimal for some i. But, consider another machine M_i' for i which has just one state in each role and never pays ρ in any period in any role. Clearly, $\|M_i^*\| > \|M_i'\|$. Moreover, by (i) above, $\pi_i(M_i', M_{-i}^*) = \pi_i(M^*) = 0$. This contradicts NEMc. $\|$

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 (part 3 of AF Theorem), in which neither player pays ρ in any period and bargaining never takes place, is also a SPEMc.[34] 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 4 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.

Proposition 5 *Fix any $\rho > 0$, and assume that M^* is a SPEMc in the costly bargaining game such that $T(M^*) < \infty$. Then, $T(M^*) = 1$.*

Proof. Suppose not. Then $\exists$ a SPEMc profile M^* that induces an agreement on some $z = (z_1, z_2) \in \triangle^2$ at some finite period $T \geq 2$. Let i be the proposer at T.

Now, given M_i^*, consider j's continuation payoff after rejecting any offer in any period at which j is the responder. Since (i) every state of M_i^* appears on the equilibrium path

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

of M^* (Lemma 1) and (ii) $\pi_j^t = \max_{f_j}(f_j, M_i^*(q_i^t))\ \forall t$ (Lemma 2), j's continuation payoff if he rejects any offer (given M_i^*) is at most $\delta^2(z_j - \rho) \geq 0$. This implies that if receives an offer $z' = (z_i + \epsilon, z_j - \epsilon) \in \triangle^2$ such that $(1-\delta^2)z_j + \delta^2\rho > \epsilon > 0$, by subgame-perfection, it must always accept.

Now, consider i using another machine $M_i' = \{M_{ip}', M_{ir}'\}$ which is identical to M_i^* except that $\lambda_{ip}'(q_i^T, (\mathcal{I}, \mathcal{I})) = z'$.

Note that $\pi_j^t(M^*) \neq \pi_j^T(M^*)\ \forall t < T$; therefore, by Lemma 3 $q_j^T \neq q_i^t\ \forall t < T$. This together with M_j^* always accepting z' implies that M_i', played against M_j^*, will generate the same outcome path up to $T-1$ as M_i^* and then agreement on z' at T. This improves i's payoff, and we have a contradiction against SPEMc. $\|$

Combining AF Theorem, Lemma 5 and Proposition 5, 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 4 *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.*

Proof. By Lemma 5 it suffices to show $T(M^*) = \infty$. To show this, suppose otherwise; then by Proposition 5 we must have that $T(M^*) = 1$. But then, by Lemma 1, M_i^* is minimal $\forall i$. Since by (part 3 of) AF Theorem perpetual disagreement is the unique stationary SPE, it follows that $T(M^*) = \infty$. This is a contradiction. $\|$

As in AF, the above selection result has been stated for the case in which the players discount the future. However, it also holds for the case of no discounting because all the proofs in this subsection are unaffected if $\delta = 1$.[35]

3.2 Costly Negotiation

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

[35]The claim in AF that perpetual disagreement is the unique SPE is also valid if there in no discounting.

the next period. (The precise details of the results here actually depend on the extensive form; whether the participation cost is paid sequentially or simultaneously. We shall discuss the latter case below.)

Let us modify some notations for the costly negotiation game. A partial history, d, now belongs to the set

$$D = \{\emptyset, (\mathcal{I}), (\mathcal{N}), (\mathcal{I},\mathcal{N}), (\mathcal{I},\mathcal{I}), (\mathcal{I},\mathcal{I},x^i), (\mathcal{I},\mathcal{I},x^i,N)\}_{i=1,2,x^i\in\triangle^2} .$$

We then have

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

The set of player i's actions is $C_i \equiv \mathcal{I} \cup \mathcal{N} \cup \triangle^2 \cup Y \cup N \cup A_i$. Also, we have

$$\begin{array}{rcl} C_{ip}(d) & = & \left\{ \begin{array}{ll} \{\mathcal{I},\mathcal{N}\} & \text{if } d = \emptyset \\ \triangle^2 & \text{if } d = (\mathcal{I},\mathcal{I}) \\ A_i & \text{if } d \in \{(\mathcal{N}), (\mathcal{I},\mathcal{N}), (\mathcal{I},\mathcal{I},x^i,N)\}_{x^i\in\triangle^2} \end{array} \right. \\ C_{ir}(d) & = & \left\{ \begin{array}{ll} \{\mathcal{I},\mathcal{N}\} & \text{if } d = (\mathcal{I}) \\ \{Y,N\} & \text{if } d = (\mathcal{I},\mathcal{I},x^j) \\ A_i & \text{if } d \in \{(\mathcal{N}), (\mathcal{I},\mathcal{N}), (\mathcal{I},\mathcal{I},x^j,N)\}_{x^j\in\triangle^2} . \end{array} \right. \end{array}$$

With this set of modifications, 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

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

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 2 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 2.

Proposition 6 *Fix any $\rho > 0$, and assume that M^* is a NEMc in the costly negotiation game such that $T(M^*) < \infty$. Then, $T(M^*) \leq 2$.*

Proof. Suppose not. Then an agreement $z \in \triangle^2$ occurs at some finite $T > 2$ under some NEMc M^*. We know from Lemma 4 that q_1^T and q_2^T are both distinct. Now let i be the proposer at T and consider three 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, we can use the arguments in Proposition 2 to derive a contradiction against NEMc.

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

There are two sub-cases to consider here.

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

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 $T > 2$ and q_j^T is distinct by Lemma 4), and (iii) $\lambda'_{jr}(q_j, (\mathcal{I})) = \mathcal{I}$ and $\lambda'_{jr}(q_j, (\mathcal{I}, \mathcal{I}, z)) = Y$.

Since 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 NEMc.

<u>*Case B2*</u>: *$\exists$ some $\tau < T$ where $\tau \in t_{ip}$ and i pays ρ.*

Consider another machine $M'_i = \{M'_{ip}, M'_{ir}\}$ which is identical to M_i^* except that (i) q_i^T is dropped, (ii) $\mu'_{ir}(q_i^{T-1}, e^{T-1}) = q_i^\tau$ (note that $q_i^\tau \neq q_i^T$ since $T > 2$ and q_i^T is distinct by Lemma 4), and (iii) $\lambda'_{ip}(q_i^\tau, (\mathcal{I}, \mathcal{I})) = z$.

Again, $M_i^{'}$ (given M_j^*) generates an identical outcome path and payoff as M_i^*. But since q_i^T is dropped, $\|M_i^*\| > \|M'_i\|$. This contradicts NEMc. $\|$

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 5 *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 $T(M^*) = \infty$.*

Proof. Part 2 follows immediately from Proposition 6. To establish part 1, by Proposition 6, it suffices to show that if $\rho > b$ then $T(M^*) \leq 2$ is not possible.

Suppose not. So, suppose that $\rho > b$ and $T(M^*) = 1$. (The other possible case in which $T(M^*) = 2$ can be treated in exactly the same way.) Let the agreement be $z \in \triangle^2$.

It then follows from Lemma 1 that M_i^* is minimal $\forall i$. Given minimality of M^* and the subgame-perfection requirement, it must be that $z_2 = (1-\delta)u_2(a) + \delta\pi_2^2(M^*)$ for some $a \in A^*$.

Now, consider player 2 using another machine $M_2' = \{M_{2p}', M_{2r}'\}$ which is identical to M_2^* except that $\lambda_{2r}'(q_2^1, (\mathcal{I})) = \mathcal{N}$.

Since M_2' is minimal, we then must have

$$\pi_2(M_1^*, M_2') = (1-\delta)u_2(a') + \delta\pi_2^2(M^*)$$

for some $a' \in A^*$. This implies that

$$\begin{aligned} \pi_2(M_1^*, M_2') - \pi_2(M^*) &= \pi_2(M_1^*, M_2') - z_2 + (1-\delta)\rho \\ &= (1-\delta)[u_2(a') - u_2(a)] + (1-\delta)\rho > 0, \end{aligned}$$

where the last inequality comes from $\rho > b$. Thus, such deviation is profitable for 2 and we have a contradiction against SPEMc. $\|$

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.

In the costly negotiation game, perpetual disagreement is not the unique equilibrium outcome if the disagreement game has multiple Nash equilibria and $0 < \rho \leq b$. This is simply because with the 2SM machine specification the disagreement game actions of a given machine in a given state can be contingent on partial histories within a period. If we consider a complexity measure based on finer partitions of histories, the result becomes sharper. Take, for instance, the 4SM specification which uses two distinct sub-machines to play bargaining (now including the participation decision) and the disagreement game within each period, and consider a corresponding SPEMc that induces immediate agreement (at $t = 1$). Then for any minimal machine profile, the associated sub-machines that play the disagreement game all have just one state and will therefore play the same (Nash equilibrium) actions whenever called upon. Thus, by the arguments behind part 1 of Theorem 5, for any $\rho > 0$ we cannot have a SPEMc inducing agreement in the first stage of the game. Together with Proposition 6 (which also holds under 4SM), this observation leads to our next Theorem. (We omit the proof and again refer the reader to Lee and Sabourian [15].)

Theorem 6 *Consider the machine game of the costly negotiation model with with 4SM specification (similar to Definition 9 above). Then, for any $\rho > 0$, every SPEMc profile M^* in 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.*

Finally, we want to mention how the analysis changes if we consider an alternative extensive form for the costly negotiation game in which the players pay the participation cost *simultaneously*.

In this case, the above selection results do not hold under lexicographic preferences ($c = 0$); in order to reproduce them we need (any) positive complexity cost ($c > 0$) and sufficiently patient players. More specifically, with sequential participation decision we first showed that if a NEMc induces an agreement then it must do so in the first two periods (Proposition 6); with simultaneous participation decision this is no longer the case if preferences are lexicographic. In such a case, a machine profile that results in an agreement at some finite $T > 2$ and no participation cost being paid (hence no bargaining) at any period before T can be a NEMc.[36] This is because neither player can deviate to a less complex machine without affecting the outcome path and payoffs. But, if $c > 0$ and δ is sufficiently high, we can show that there is a profitable deviating opportunity for a player who trades off a small loss in payoff for a saving in complexity cost. This allows us to establish the results of Proposition 6 for the case of $c > 0$ and sufficiently high δ.

Proposition 7 *Consider the costly negotiation game with simultaneous participation decision. Fix any $c > 0$ and any $\rho > 0$. Then, $\exists\ \bar{\delta} < 1$ such that, for any $\delta \in (\bar{\delta}, 1)$, if M^* is a NEMc such that $T(M^*) < \infty$ we have $T(M^*) \leq 2$.*

[36] In such an equilibrium, player i's machine would have T states $\{q_i^1, .., q_i^T\}$ with initial state q_i^1, and output and transition functions satisfying respectively

$$\lambda_i(q_i^t, \emptyset) = \begin{cases} \mathcal{N} & \text{if } t < T \\ \mathcal{I} & \text{if } t = T \end{cases}$$

$$\mu_i(q_i^t, e) = \begin{cases} q_i^{t+1} & \text{if } e = (\mathcal{N}, \mathcal{N}) \text{ and } t < T \\ q_i^1 & \text{otherwise} \end{cases}$$

Proof. See Appendix. ‖

By the same reasoning as in Theorem 5 and its Corollaries, Proposition 7 enables us to deliver the same conclusion for the case of simultaneous participation decision with $c > 0$ and sufficiently high δ.

Theorem 7 *Consider the costly negotiation game with simultaneous participation decision. Fix any $c > 0$ and any $\rho > b$. Then $\exists\ \bar{\delta} < 1$ such that, for any $\delta \in (\bar{\delta}, 1)$, every SPEMc profile M^* 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.*[37]

4 Conclusion

By considering players who care for complexity of a strategy (at the margin) as well as payoffs, we can provide very sharp predictions in several games of negotiation which otherwise admit a large number of equilibria.

The nature of the selected equilibria, however, differs markedly in these games. While the negotiation game only admits equilibria (SPEMc) that are *efficient*, the only SPEMc outcome in the costly bargaining game is its most *in*efficient one that involves perpetual disagreement. Moreover, we find that introducing small transaction costs to the negotiation game can swing the selection results sharply towards inefficiency: the only SPEMc outcome in the costly negotiation game is perpetual disagreement, which is of course inefficient if every disagreement game outcome is dominated by an agreement.

We draw the following conclusions. First, the efficiency property of the selected equilibria in the negotiation game suggests that the Coase Theorem can be indeed extended to negotiation models in which disagreement payoffs are endogenously determined. In other words, the theorem is valid even in repeated game contexts. We thus complement the existing repeated game literature in which considering complexity or bargaining alone has produced only a limited selection result. Second, the selection results in the costly bargaining and negotiation games demonstrate that transaction costs are a critical ingredient of a robust account of why inefficient outcomes can arise in two-person negotiation even with complete information. It is particularly interesting to observe that only an infinitesimal amount of transaction cost can shift the SPEMc outcome so dramatically from efficiency to inefficiency in the negotiation game.

[37] As with the sequential model (Theorem 6), the assumption $\rho > b$ in Theorem 7 can be replaced by $\rho > 0$ if G has a unique Nash equilibrium and/or we adopt 4SM.

5 Appendix: Relegated Proofs

Proof of Proposition 1. This proof is a direct extension of the proof of Theorem 1 in KS. Let $M_i = \{M_{ip}, M_{ir}\}$ be any machine that implements some strategy f_i where $M_{ik} = (Q_{ik}, q^1_{ik}, \lambda_{ik}, \mu_{ik})$ for $k = p, r$.

First, we show that $|Q_{ik}| \geq |F_{ik}(f_i)|\ \forall k$. For any $q_i \in Q_{ik}$ and $k = p, r$, let $M_i(q_i) = \{M_{ik}(q_i), M_{il}\}$ be the machine that is identical to M_i except that it starts with the sub-machine $M_{ik}(q_i)$ where $M_{ik}(q_i) = (Q_{ik}, q_i, \lambda_{ik}, \mu_{ik})$ is the sub-machine identical to M_{ik} except for the initial state q_i.

Note that for every $\bar{f}_i \in F_{ik}(f_i)$ and $k = p, r$, there exists some $h \in H^\infty_{ik}$ such that $\bar{f}_i = f_i|h$. Now define a function $\Gamma_{ik} : Q_{ik} \to F_{ik}(f_i)$ such that $\Gamma_{ik}(\bar{q}_i)$ is the strategy implemented by $M_i(\bar{q}_i)$ for any $\bar{q}_i \in Q_{ik}$. It then follows that for every $\bar{f}_i \in F_{ik}(f_i)$, there must exist a distinct state $\bar{q}_i \in Q_{ik}$ such that $\Gamma_{ik}(\bar{q}_i) = \bar{f}_i$. Simply let $\bar{q}_i = q_i(h)$ (as defined inductively in Section 3) where h is the history such that $\bar{f}_i = f_i|h$.[38]

Second, we show that there exists a machine implementation of f_i which only uses $F_{ip}(f_i)$ and $F_{ir}(f_i)$ as the set of states for its corresponding sub-machines.

Define $\bar{M}_i = \{\bar{M}_{ip}, \bar{M}_{ir}\}$ such that, for $k = p, r$, $\bar{M}_{ik} = (F_{ik}(f_i), f^1_{ik}, \bar{\lambda}_{ik}, \bar{\mu}_{ik})$ where

- $f^1_{ik} \in F_{ik}(f_i)$ is the initial state and if i plays role k at the initial history then $f^1_{ik} = f_i$
- for any $\bar{f}_i \in F_{ik}(f_i)$ and $d \in D_{ik}$, $\bar{\lambda}_{ik}(\bar{f}_i, d) = \bar{f}_i(\emptyset, d)$ where $\emptyset$ is the empty history
- $\bar{\mu}_{ik}(\bar{f}_i, e) = \bar{f}_i|h, e$ for any $h \in H^\infty_{ik}$ and $e \in E$.

This machine has $\sum_k |F_{ik}(f_i)|$ states (each k sub-machine with $|F_{ik}(f_i)|$ states) and implements f_i. ‖

Proof of Proposition 3. (i) Necessity follows from the property that in any stationary SPE the players must play a Nash equilibrium of G after any rejection. Sufficiency follows from Corollary 1 of BW.

(ii) Suppose not. Then f is a stationary SPE, and for some $\epsilon > 0$,

$$\sum_i \pi_i(f) + \epsilon < 1 - (1 - \delta)b \ . \tag{12}$$

Next, consider player 1 making a deviating offer $z = (z_1, z_2) \in \triangle^2$ at $t = 1$ such that

$$z_2 = (1 - \delta)(u_2(a^1) + b) + \delta\pi^2_2(f) + \epsilon = \pi_2(f) + (1 - \delta)b + \epsilon \tag{13}$$

[38]It is critical here that each sub-machine uses its own distinct set of states. Otherwise, a single state can be used to activate two distinct continuation strategies, one in each role.

where a^1 is the disagreement game equilibrium outcome in $t = 1$ when f is chosen. Clearly, given subgame-perfection of the stationary profile f and the definition of b, f_2 will accept such offer. Also, by subgame-perfection such deviation must be unprofitable for player 1; therefore $z_1 \leq \pi_1(f)$. This, together with (13), implies that

$$1 - \pi_2(f) - (1-\delta)b - \epsilon \leq \pi_1(f) \ .$$

But this contradicts (12). ‖

Proof of Corollary 3. (i) If G has a unique Nash equilibrium, $b = 0$. Thus, the claim follows from Proposition 3.

(ii) Let f be a stationary SPE and assume it is inefficient. There are two possible cases to consider.

<u>Case A</u>: *There is one period of delay, followed by an agreement.* We know from Proposition 3 that $\sum_i \pi_i(f) \geq 1 - (1-\delta)b$. Thus, it must be that

$$(1-\delta)\sum_i u_i(a) + \delta \geq 1 - (1-\delta)b$$

where $a \in A^*$ is the disagreement game outcome in the first period. But this is clearly not possible if $\forall a \in A^*$ $\sum_i u_i(a) < 1 - b$ as in the claim.

<u>Case B</u>: *There is infinite delay.* Let $a^1 \in A^*$ and $a^2 \in A^*$ be the disagreement outcome in odd and even periods respectively. Then, by Proposition 3, we must have

$$\frac{\sum_i u_i(a^1) + \delta \sum_i u_i(a^2)}{1+\delta} \geq 1 - (1-\delta)b \ .$$

Again, this is not possible if $\sum_i u_i(a) < 1 - b$, $\forall a \in A^*$. ‖

Proof of Theorem 2. Fix any $c \in (0,1)$.[39] Define $\bar{\delta} = 1 - \frac{c}{\beta}$ where β is given by (4) above. By Corollary 2, it suffices to establish that for any $\delta \in (\bar{\delta}, 1)$ every SPEMc is stationary. Suppose not; then there exist a discount factor $\delta \in (\bar{\delta}, 1)$ and a profile $M^* = (M_1^*, M_2^*) \in \Omega^\delta(c)$ such that both equilibrium machines M_1^* and M_2^* are non-minimal. (Clearly, if one equilibrium machine is minimal then the other equilibrium machine also has to be minimal.) Also, by Proposition 2, it must be that $T(M^*) = \infty$.

Similarly to the proof of Proposition 4 above, define η such that

$$0 < \eta < c - \beta(1-\delta) \ . \qquad (14)$$

[39] The case of $c \geq 1$ is trivial because then complexity cost (weakly) dominates any feasible average payoff for each player in the negotiation game and thus any equilibrium machine must be minimal. We can refer to BW Result 1 for SPEMc characterization in this case.

Define as before t_{ik}, τ_η, and τ (see (6), (7), and (9)).

We need to consider the following two cases separately.

Case A: $\tau_\eta > 2$.

First note that

$$q_1^t \neq q_1^{\tau_\eta} \quad \forall t < \tau_\eta \ . \tag{15}$$

Otherwise, $q_1^t = q_1^{\tau_\eta}$ for some $t < \tau_\eta$. But then, $\pi_2^t = \pi_2^{\tau_\eta}$ by Lemma 3. This contradicts the definition of τ_η.

Next, consider player 1 using another machine $M_1' = \{M_{1p}', M_{1r}'\}$ which is identical to M_1^* except that (i) $q_1^{\tau_\eta}$ is dropped and (ii) $\lambda_{1p}'(q_1^{\tau_\eta - 1}, \emptyset) = (1 - \pi_{2r}^{\max}, \pi_{2r}^{\max}) = z'$, where $\pi_{2r}^{\max}$ is defined by (8) with η now given by (14).

By (15), dropping $q_1^{\tau_\eta}$ does not affect the outcome path up to τ. Therefore, by similar arguments as in the proof of Proposition 4, the deviation would induce agreement z' in period τ. It then follows that 1's deviation payoff here is given also by $1 - \pi_{2r}^{\max}$, and as in (11), we have

$$1 - \pi_{2r}^{\max} \geq 1 - (\pi_2^\tau + (1-\delta)\beta + \delta\eta) \ .$$

We also have that $1 - \pi_2^\tau \geq \pi_1^\tau$. Thus, 1's loss from such deviation is

$$\pi_1^\tau - (1 - \pi_{2r}^{\max}) \leq (1-\delta)\beta + \delta\eta \ .$$

But M_1' has one less state than M_1^* (since $q_1^{\tau_\eta}$ has been dropped) implying that the deviation also results in a saving on complexity cost by $c > 0$. Since $c > (1-\delta)\beta + \eta$ by (14), we have

$$\pi_1^\tau - (1 - \pi_{2r}^{\max}) < c \ ,$$

and therefore, the loss from deviation is less than c. Thus, M_1' is a *c-best response* to M_2^* and has one less state. This contradicts SPEMc.

Case B: $\tau_\eta = 2$.

We know that M_1^* is not minimal. Then, consider player 1 using another minimal machine $M_1' = \{M_{1p}', M_{1r}'\}$ which is constructed such that $Q_{1p}' = \{q_1^1\}$, $Q_{1r}' = \{q_1^2\}$ and $\lambda_{1p}'(q_1^1, \emptyset) = z'$ (where z' is as appearing in Case A above).

By similar arguments to those behind Case A above, such deviation induces immediate agreement on z' and is (overall) profitable for 1 since M_1' is minimal and has less states than M_1^*. Thus we again have a contradiction against SPEMc. ‖

Proof of Proposition 7. Suppose not. Let β be as defined by (4) above, and define

$$\bar{\delta} = 1 - \frac{c}{\beta + \rho} < 1 \ .$$

Then $\exists\ \delta \in (\bar{\delta}, 1)$ and a profile M^* that is a NEMc in the game with discount factor δ such that $T(M^*) \in (2, \infty)$. Let i be the proposer at T. Also, define

$$\tau = \min\{t \in t_{ir} \mid \pi_i^t \leq \pi_i^{t'} \ \forall t' \in t_{ir}\} \ .$$

We know that:

- q_i^t is distinct for $\forall t \leq T$ and $\forall i$ (Lemma 4)
- neither player pays ρ at any $t \in t_{ip}$ where $t < T$ (otherwise, use similar arguments behind Proposition 6 to show a contradiction).

Now, consider another machine $M_i' = \{M_{ip}', M_{ir}'\}$ which is identical to M_i^* except that (i) $q_i^{\tau-1}$ is dropped and replaced by q_i^T (thus $\mu_{ir}'(q_i^{\tau-2}, e^{\tau-2}) = q_i^T$) and (ii) transition from q_i^T is such that $\mu_{ip}'(q_i^T, \tilde{e}) = q_i^{\tau'}$, where

$$\tilde{e} = \big(\mathcal{I}, \mathcal{N}, \lambda_i^*(q_i^T, (\mathcal{I}, \mathcal{N})), \lambda_j^*(q_j^{\tau-1}, (\mathcal{I}, \mathcal{N}))\big) \in E$$

is the outcome generated by q_i^T and $q_j^{\tau-1}$) and $\tau' \in t_{ir}$ is such that $\mu_j^*(q_j^{\tau-1}, \tilde{e}) = q_j^{\tau'}$.

Since every q_i^t ($t \leq T$) is distinct, playing M_i' against M_j^* replicates the outcome path up to $\tau - 2$ and then generate a deviation at $\tau - 1$ (where i now pays ρ). By Lemma 1 and the definition of τ, this implies that

$$\pi(M_i', M_j^*) \geq (1-\delta)\left[\sum_{t=1}^{\tau-2} \delta^{t-1} u_i(a^t) - \rho_i^{\tau-2} + \delta^{\tau-2}(\inf_a u_i(a) - \rho)\right] + \delta^{\tau-1}\pi_i^{\tau}.$$

Since

$$\pi_i(M^*) < (1-\delta)\left[\sum_{t=1}^{\tau-2} \delta^{t-1} u_i(a^t) - \rho_i^{\tau-2} + \delta^{\tau-2} \sup_a u_i(a)\right] + \delta^{\tau-1}\pi_i^{\tau} \ ,$$

it then follows that

$$\begin{aligned} \pi_i(M^*) - \pi_i(M_i', M_j^*) &\leq (1-\delta)\delta^{\tau-2}\left[\sup_{a,a'}(u_i(a) - u_i(a')) + \rho\right] \\ &< (1-\delta)\left[\sup_{a,a'}(u_i(a) - u_i(a')) + \rho\right] \ , \end{aligned}$$

where the second inequality follows from $\delta < 1$.

Thus, playing M_i' is overall profitable for i (given $c > 0$) if

$$c > (1-\delta)\left[\sup_{a,a'}(u_i(a) - u_i(a')) + \rho\right]$$

which is indeed the case given $\delta > \bar{\delta}$. This contradicts NEMc. $\|$

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] Binmore, K., M. Piccione, and L. Samuelson, Evolutionary Stability in Alternating-Offers Bargaining Games, *Journal of Economic Theory*, 80 (1998), 257-291.

[5] Bloise, G., Strategic Complexity and Equilibrium in Repeated Games, unpublished doctoral dissertation, University of Cambridge, 1998.

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

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

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

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

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

[11] Fershtman, C., and D. J. Seidmann, Deadline Effects and Inefficient Delay in Bargaining with Endogenous Commitment, *Journal of Economic Theory*, 60 (1993), 306-321.

[12] Gale, D., and H. Sabourian, Complexity and Competition, forthcoming *Econometrica.*

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

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

[15] Lee, J., and H. Sabourian, Complexity and Efficiency in Repeated Games with Negotiation, *Cambridge Working Papers in Economics*, 0419 (2004).

[16] Nash, J., Two-Person Cooperative Games, *Econometrica*, 21 (1953), 128-140.

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

[18] Perry, M., and P. J. Reny, A Non-cooperative Bargaining Model with Strategically Timed Offers, *Journal of Economic Theory*, 59 (1993), 50-77.

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

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

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

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

[23] Sákovics, J., Delay in Bargaining Games with Complete Information, *Journal of Economic Theory*, 59 (1993), 78-95.