Hitotsubashi University
Department of Economics

2-1 Naka, Kunitachi, Tokyo, Japan

Discussion Paper #2013-15

A Non-cooperative Bargaining Theory with Incomplete Information: Verifiable Types

Akira Okada

December 2013
October 2014 (revised)

A Non-cooperative Bargaining Theory with Incomplete Information: Verifiable Types*

Akira Okada†

December 2013
October 2014 (revised)

Abstract

We consider a non-cooperative sequential bargaining game with incomplete information where two players negotiate for mechanisms with ex post verifiable types at the interim stage. We prove the existence of a stationary sequential equilibrium of the bargaining game where the ex post Nash bargaining solution with no delay is asymptotically implemented with probability one. Further, the ex post Nash bargaining solution is a unique outcome of a stationary equilibrium under the property of Independence of Irrelevant Types (IIT), whereby the response of every type of a player is independent of allocations proposed to his other types, and under a self-selection property of their belief. Interim efficiency (insurance benefit) in the Bayesian bargaining problem is not necessarily supported in a non-cooperative approach.

JEL classification: C72; C78; D82

Keywords: Bayesian bargaining problem; incomplete information; mechanism selection; ex post Nash bargaining solution; non-cooperative games

*I am grateful to seminar participants at Hitotsubashi University, Hokkaido University, SAET2014, University of Cambridge, University of Tokyo and Waseda University for their valuable comments. Financial support from the Japan Society for the Promotion of Science under grant No.(A)26245024 is gratefully acknowledged.

†Graduate School of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan. Phone: +81 42 580 8599, Fax: +81 42 580 8748, E-mail: aokada@econ.hit-u.ac.jp

1 Introduction

We consider a two-person bargaining problem with incomplete information in which each player has private information about his type. Knowing their own types, players negotiate for a contract (or a mechanism) that is a contingency plan that prescribes a joint action for every possible type profile of players. Players' private information may affect their preferences over agreements. To reach a preferable agreement, players may want to reveal or conceal their types. Private information may leak through actions in negotiations. A bargaining situation is called a case of *verifiable types* if players' types become publicly known and verifiable when an agreement is implemented. In the other case of *unverifiable types*, a contract should satisfy the Bayesian incentive compatibility so that players have incentives to disclose their types truthfully.[1] To focus on the analysis of bargaining behavior with incomplete information, we assume the condition of verifiable types in this paper.

As an example, consider disarmament negotiations between two countries (Harsanyi and Selten 1972). Neither country has precise knowledge about the other country's armament levels, technology, political and economic conditions, and utility values to possible agreements. These uncertain variables are represented by players' types in games with incomplete information. A physical action prescribed by a disarmament treaty may or may not depend on types of countries. It may simply require the countries to destroy an absolute number of military weapons by each side. In a general case, it may require countries to reduce the numbers of weapons in their possession, depending on their types. For example, some agreements require each country to destroy a certain percentage of its total stockpile of missiles. Others may prescribe the number of missiles destructed by either side according to a mathematical function of both sides' missile stocks.[2] Such type-dependent agreements can be implemented

[1]When players' types represent their internal states such as satisfaction, risk attitudes and psychological characters, it is appropriate for us to model them as unverifiable types.

[2]The new START treaty agreed by the United States and Russia on April 8, 2010 pre-

only if countries' types are truthfully disclosed. To overcome the enforcement problem, a disarmament treaty involves a specified procedure which provides both countries opportunities of inspection and verification about each other's type to the extent that the treaties can be enforced.[3]

In this paper, we consider a Rubinstein (1982)-type sequential bargaining process under incomplete information. Knowing his own type, a randomly selected player proposes a contract. The other player either accepts or rejects it. If he accepts it, then the contract is agreed upon. Thereafter, a process of verification is conducted, and an action prescribed by the contract for both players' types is jointly taken. If the proposal is rejected, then there is the risk that negotiations may fail with a positive probability. In this case, a predetermined outcome results. In the example of disarmament negotiations, the status-quo prevails. If negotiations may not fail (with the remaining probability), then negotiations go to the next round, and the same process is repeated over (possibly) infinitely many rounds until an agreement is made.

Our bargaining model with verifiable types can be applied to some economic situations under uncertainty. For example, consider an insurance contract which prescribes how much a customer should be covered for contingencies by an insurer. At the time of trading, both parties have only imperfect and private information about which event may happen. When the contract is implemented, the insurer is assumed to have a sufficient ability to verify a realized damage. Sharecropping is another example. A tenant and a landlord negotiate for a rental share contract that specifies the proportion of outputs the tenant should deliver to the landlord. The outputs may depend on un-

scribes that each party shall reduce and limit its armaments so that the aggregate numbers are (a) 700 for deployed ICBMs, deployed SLBM, and deployed heavy bombers, (b) 1550 for warheads on deployed ICBMs, warheads on deployed SLBMs, and nuclear warheads counted for deployed heavy bombers, and (c) 800 for deployed and non-deployed ICBM launchers, deployed and non-deployed SLBM launcher, and deployed and non-deployed heavy bombers (Article II).

[3]The new START treaty involves various measures to ensure compliance: creation and notification of database (Article VII), exchange of telemetric information (Article IX), national technical means of verification (Article X) and inspections (Article XI).

certain events such as weather and other agricultural conditions. While the output is publicly known and verifiable to both parties at the time of contract implementation, they are only partly informed about it at the time of contracting.

The result of the paper is as follows. We consider the ex post Nash bargaining solution of the two-person Bayesian bargaining problem, which is a specific contract that assigns the Nash bargaining solution to every type profile of players. We first show that there exists a sequential equilibrium of the bargaining game which implements the ex post Nash bargaining solution in the limit that the continuation probability of negotiations in case of rejection (alternatively, the discount factor of future payoffs) goes to one. The constructed equilibrium satisfies several properties: (stationarity) every player's equilibrium strategy depends only on his own types, independent of past actions; (no delay of agreement) an agreement is made with probability one in the initial round; (inscrutability) every type of a proposer proposes the same contract; (information revealing) a proposer may update his prior belief about a proposer based on revealed information when he receives an unexpected proposal.

In the last part of the paper, we provide a characterization result of a sequential equilibrium of the bargaining game satisfying the properties above. We prove that the ex post Nash bargaining solution is asymptotically a unique outcome of a stationary sequential equilibrium satisfying the property of independence of irrelevant types (IIT) and a refinement condition based on self-selection. IIT means that the response of every type of a player depends only on a proposal made to himself, independent of allocations proposed to his other (irrelevant) types. We prove that no delay of agreement occurs in a stationary equilibrium with IIT. Our refinement concept of a sequential equilibrium, similar to the notion of a perfect sequential equilibrium of Grossman and Perry (1986a), assumes that, if a responder is offered an unexpected proposal, then he infers that a true type of a proposer must be among those who have incentives to make the proposal, and that he updates his prior belief based on the revealed information.

The result of the paper has somewhat a surprising implication. Since the ex post Nash bargaining solution may not be interim efficient, our non-cooperative approach to the Bayesian bargaining problem does not support any cooperative solution assuming interim efficiency, for example, those obtained in the axiomatic approach of Harsanyi and Selten (1972) and Myerson (1984). Relating to this, the result implies that insurance benefit based only on private information is impossible because players' private information may be revealed in the process of negotiations.

The literature on the Bayesian bargaining games with incomplete information is diverse. In their pioneering work, Harsanyi and Selten (1972) extend the bargaining theory of Nash (1950) to the case of incomplete information with verifiable types. They consider a non-cooperative multi-stage model of bargaining. To select a unique equilibrium of the bargaining model, they develop an axiomatic theory based on Nash (1950) and present a generalized Nash solution (called the Harsanyi–Selten solution) under incomplete information. Myerson (1979) applies the Harsanyi–Selten solution to the case of unverifiable types in which incentive compatibility is required for a feasible agreement. In a subsequent paper, Myerson (1984) acknowledges a theoretical drawback of the Harsanyi–Selten solution in that it violates a decision-theoretic axiom of probability-invariance. He considers an alternative set of axioms and defines a set-valued solution called a neutral bargaining solution as the minimal solution satisfying his axioms. The Myerson solution coincides with the classical Nash bargaining solution when it is applied to the bargaining game with complete information.

Since the work of Harsanyi and Selten (1972), non-cooperative analysis of the Bayesian bargaining problem has been mainly done for the principal-agent set-up in which a principal has all the bargaining power. Most studies are restricted to the ultimatum bargaining model in which a principal makes a take-it-or-leave-it offer of a contract to an agent (or agents). There is a large volume of works on an uninformed principal in the literature of adverse-selection (or screening) models. Remarkably, Myerson (1983) considers the mechanism design problem of an informed principal. To deal with the multiplicity of sequential equilibria, he applies a cooperative axiom in the core theory and presents a set-valued solution. Maskin and Tirole (1990, 1992) elaborate a non-cooperative analysis of the informed principal model and characterize a perfect Bayesian equilibrium for two cases of private and common values. de Clippel and Minelli (2004) refine Myerson's work in the case of verifiable types. Mylovanov and Tröger (2012) extend the result of Maskin and Tirole (1990) to a general case of private values. To our best knowledge, there are few works on sequential bargaining games for mechanism selection.[4]

The remainder of the paper is organized as follows. Section 2 presents the model. Section 3 proves the existence of a sequential equilibrium which implements the ex post Nash bargaining solution asymptotically. Section 4 gives the no-delay result under IIT. Section 5 provides a characterization result. Section 6 discusses the result of the paper. Section 7 concludes. Some proofs are given in Appendix.

2 The Model

We consider a two-person bargaining problem with incomplete information, following Myerson (1984). Let $N = \{1, 2\}$ be the set of players. For each $i = 1, 2$, let T_i be a finite set of player i's types t_i. Let $T = T_1 \times T_2$. An element of T is denoted by $t = (t_1, t_2)$. For each $t_i \in T_i$, $T(t_i)$ denotes the cylinder set $\{t_i\} \times T_j (j \neq i)$. Let π, a probability distribution on T, denote the

[4]There exists another branch of the literature which considers various sequential bargaining games with incomplete information. In these games, players with private information propose type-independent allocations. Typical observations are that there is a large set of sequential equilibria, and that the equilibrium delay of an agreement may happen. The literature includes Fudenberg and Tirole (1983), Rubinstein (1985), Grossman and Perry (1986b) and Chatterjee and Samuelson (1987) among others. Ausubel et al. (2002) present an excellent review on the literature.

common *prior* belief of players. For each $t \in T$, $\pi(t)$ denotes the probability that type profile t is realized. We assume that $\pi(t) > 0$ for all $t \in T$. For each $t_i \in T_i$, the *posterior* belief of player i given t_i is defined by

$$\pi(t_j|t_i) = \frac{\pi(t_i, t_j)}{\sum_{t'_j \in T_j} \pi(t_i, t'_j)}. \tag{1}$$

Let A be the set of actions (or outcomes) available to players if they cooperate. A specific element $d^* \in A$ is called the *disagreement action*, and describes the action that prevails when cooperation fails. We assume that the disagreement action d^* is exogenously given. For each $i = 1, 2$, the function $u_i : A \times T \to R$ denotes a state-dependent von Neumann–Morgenstern utility function for player i. Without loss of generality, we normalize u_i so that $u_i(d^*, t) = 0$ for all $t \in T$ and all $i = 1, 2$.

A *two-person Bayesian bargaining game* is represented by $G = (A, d^*, T_1, T_2, u_1, u_2, \pi)$. For each $t \in T$, let $U(t)$ denote the set of payoff vectors $u(a, t) = (u_i(a, t))_{i=1,2}$ of players for all actions $a \in A$. A payoff vector $u = (u_1, u_2)$ of $U(t)$ is *Pareto efficient* if there is no other $v = (v_1, v_2) \in U(t)$ such that $v_i \geq u_i$ for all $i = 1, 2$ and $v_i > u_i$ for some i. The *Pareto frontier* of $U(t)$ is the set of all Pareto efficient payoff vectors of $U(t)$. We represent the Pareto frontier of $U(t)$ as an equation $H^t(u_1, u_2) = 0$, and call H^t the Pareto frontier function of $U(t)$. Without loss of generality, we can assume that $H^t(u_1, u_2) \geq 0$ for all $u \in U(t)$. We make the following assumptions.

Assumption 2.1. For every $t \in T$,

(1) $U(t)$ is a nonempty, convex and compact subset of R^2,

(2) the Pareto frontier of $U(t)$ intersects the two axes, $u_1 = 0$ and $u_2 = 0$, of R^2, and

(3) the disagreement payoff $u(d^*, t) = (0, 0)$ is an interior point of $U(t)$.

These assumptions of the feasible set $U(t)$ are standard in the literature.

For each $t \in T$, a payoff vector $u = (u_1, u_2)$ of $U(t)$ is *individually rational* if $u_i \geq 0$ for all $i = 1, 2$. $U_+(t)$ denotes the set of all individually rational payoffs of $U(t)$. Assumption 2.1 guarantees the existence of the implicit function $u_i = h_i^t(u_j)$ satisfying $H^t(h_i^t(u_j), u_j) = 0$ on $U_+(t)$ for every $i, j = 1, 2$ and $i \neq j$.

In the game G, players negotiate for a mechanism, not for a single action in A. A *mechanism* x is a contract specifying which action should be chosen jointly, contingent on the player types. Formally, x is a function from T to A. Let M be the set of all mechanisms. Under a mechanism x, players are supposed to choose an action $a = x(t)$ when they are of type t. When a mechanism x is implemented, the conditional expected utility $Eu_i(x|t_i)$ of player i given type t_i is defined by

$$Eu_i(x|t_i) = \sum_{t_j \in T_j} \pi(t_j|t_i) u_i(x(t), t). \tag{2}$$

As we have noted in the introduction, we assume that players' types become publicly known and verifiable when a mechanism is implemented, although they are privately known to players during negotiations. In this case of verifiable types, any mechanism is implementable for the players as long as it is physically feasible. The Bayesian incentive compatibility is irrelevant in this paper.

We are interested in a specific mechanism where two players choose the Nash bargaining solution for every type profile $t \in T$.

Definition 2.1. A mechanism x^{NB} in G is called the *ex post Nash bargaining solution* with weights $p = (p_1, p_2)$ if it assigns to every $t \in T$ the Nash bargaining solution $x^{NB}(t)$ of the feasible set $U(t)$ with weights $p = (p_1, p_2)$. The payoff vector $(u_1(x^{NB}(t), t), u_2(x^{NB}(t), t))$ maximizes the Nash product $u_1^{p_1} \cdot u_2^{p_2}$ over $U(t)$ for every $t \in T$, where the disagreement payoff is given by $u(d^*, t) = (0, 0)$.

We formulate a negotiation process over mechanisms with incomplete information as a sequential bargaining game in the spirit of Rubinstein (1982). Specifically, as a bargaining protocol, we apply the random proposer rule, which has been well studied in the literature on non-cooperative bargaining games with complete information (Binmore, Rubinstein and Wolinsky 1986, Baron and Ferejohn 1989, and Okada 1996 among others).

Negotiations take place at an *interim* stage, in which players know their own type but not that of the other player. After the player types are realized and revealed to them privately, one player is randomly selected as a proposer and proposes a feasible mechanism to the other player. If the opponent accepts it, then the proposed mechanism is agreed. Any agreed-upon mechanism will be implemented at an *ex post* stage where players' types become publicly known.[5] Otherwise, negotiations may stop with probability $\epsilon > 0$, and the disagreement action d^* is chosen. With probability $1 - \epsilon$, negotiations may continue in the next round. If this happens, then a new proposer is randomly selected again, and the same process is repeated. The probability of an infinite play in negotiations is zero. Players are assumed to maximize their expected payoffs.

An alternative interpretation of the negotiation model is that negotiations continue in the next round in the case of rejection, and that players discount their future payoffs by $\delta = 1 - \epsilon$. The disagreement action d^* prevails in the case of no agreements. For sake of exposition, we will employ this interpretation of the model with a discount factor $\delta < 1$ in what follows.

Formally, the bargaining game has two stages of negotiations and of implementation. The first stage of negotiations has the following rule. In round 0, a type profile $t = (t_1, t_2)$ of players is realized according to the prior prob-

[5]This modelling assumption does not mean that players' types become verifiable upon agreement to a mechanism. In the example of disarmament negotiations, two countries engage in mutual inspection and verification according to an agreement before a treaty is implemented. In a case of sharecropping, an uncertain amount of crops becomes publicly known in a harvesting season when a contract is implemented.

ability distribution π. Every player $i(=1,2)$ knows his own type t_i, but not that of the other player, t_j. At the beginning of round 1, a player is randomly selected as a proposer according to a predetermined probability distribution $p = (p_1, p_2)$. The selected player proposes a feasible mechanism $x \in M$. The other player either accepts or rejects the proposal. If the responder accepts the proposal, then x is agreed. If not, then the game continues in round 2, and a new proposer is randomly selected. The same process as above is repeated until some mechanism is agreed. The negotiation stops if an agreement is reached, and thereafter the agreement is implemented in the second stage where players' types become publicly known. When the negotiation does not stop, the disagreement outcome d^* prevails in the implementation stage.[6] Let $\delta(<1)$ be the common discount factor for future payoffs of players.

We denote by Γ^δ the bargaining game with incomplete information introduced above. Whenever each player makes a choice in Γ^δ, he knows perfectly his own type and all past moves, including the selection of proposers. However, a player does not know the other player's type. We sometimes omit δ in the notation Γ^δ if no confusion arises.

A strategy for every player in Γ is defined in a standard way. A *(pure) strategy* σ_i for player i in Γ is a function that assigns a choice to each of his possible moves, depending on the information he receives. Specifically, σ_i prescribes a mechanism $\sigma_i(t_i, h) \in M$ in every round when player i is a proposer, given his type t_i and a history h of play before the round. In addition, σ_i prescribes a response $\sigma_i(t_i, h, x) \in \{accept, reject\}$ to every proposal x when he is a responder. For a strategy profile $\sigma = (\sigma_1, \sigma_2)$, the expected (discounted) utility $Eu_i(\sigma)$ for each player i is defined in a standard way.

[6]Our model has the time structure that the disagreement action d^* is played after infinitely many bargaining rounds. It is constructed to describe bargaining situations without any "end effect." It is assumed that players commonly perceive that there would be a chance that negotiations continue in the next round when a proposal is rejected. As long as a stopping probability is positive in each bargaining round, the probability of an infinite number of bargaining rounds is zero. In real disarmament negotiations, it seems reasonable to assume that there would be a chance, however small, that countries may continue negotiations even after a proposal is rejected.

A *belief* system for Γ is a function μ that assigns every player i his belief about the other player's type, a probability distribution on T_j. Given (t_i, h), let $\mu(t_j|t_i, h)$ denote the belief of player i about t_j when he is a proposer, and let $\mu(t_j|t_i, h, x)$ be his belief when he responds to a proposal x from player j.

We employ a sequential equilibrium (Kreps and Wilson 1982) as a non-cooperative solution concept for the bargaining game Γ. Roughly, a pair (σ, μ) of a strategy profile and a belief system is a sequential equilibrium of Γ if the strategy of every player is a best response to the other's strategy for each of his information sets under the belief system μ, where μ should be consistent with the strategy profile σ (and with some slight deviation from it off equilibrium play) by the Bayes rule. Since the notion of a sequential equilibrium is standard, we omit a precise definition.

The multiplicity of a sequential equilibrium is a central issue of the sequential bargaining theory. Rubinstein (1982) shows that his two-person sequential bargaining game with complete information has a unique subgame perfect equilibrium, which is composed of stationary (history-independent) strategies. The uniqueness of a subgame perfect equilibrium does not hold if $n \geq 3$ (see Sutton 1986 and Osborne and Rubinstein 1990). In the case of incomplete information, Rubinstein (1985) shows that the set of sequential equilibria is very large even in the two-person case, due to the freedom of players' constructing beliefs off the equilibrium play.

In this paper, we consider a stationary equilibrium of the bargaining game Γ with incomplete information. The definition of a stationary equilibrium is as follows.

Definition 2.2. A sequential equilibrium (σ, μ) of Γ is said to be *stationary* if every player i's behavior in every round depends only on his type t_i: specifically, (i) a proposer's behavior depends only on his type, and (ii) a responder's behavior depends only on his type and a proposal.

This definition of a stationary equilibrium for sequential bargaining games with incomplete information is essentially the same as that with complete information in the literature. That is, players' proposals and responses in each round are independent of past actions. A usual justification for a stationary equilibrium is a focal-point (or reference point) argument.[7] It is the simplest form of bargaining strategies, and it may be easier for bargainers to coordinate their expectations.

We remark that some type of "learning" may happen in a stationary equilibrium. In particular, a central issue in this paper is what a responder may learn about a type of a proposer when he receives an unexpected proposal off the equilibrium play. The notion of a sequential equilibrium, however, is not sufficient to the study of this issue since it allows an arbitrary belief of the responder off the equilibrium play. We consider a situation where a proposal may reveal some information of a proposer's type if he has an incentive to screen himself. The responder may update his belief based on such revealed information.

In what follows, we refer to a stationary sequential equilibrium simply as a stationary equilibrium. For a stationary equilibrium σ and a type profile $t \in T$, we denote by $Eu_i(\sigma|t)$ the conditional expected (discounted) payoff of player i for σ evaluated at the beginning of each bargaining round before the random selection of a proposer occurs. Since σ is stationary, $Eu_i(\sigma|t)$ is independent of past actions. Whenever no confusion arises, we use a simpler notation $v_i(t)$ for $Eu_i(\sigma|t)$.

[7] There are divergent views among researchers about whether a stationary equilibrium (a Markov-perfect equilibrium, in general) is a reasonable solution for sequential bargaining games. For a positive theory of bargaining, it is an important question whether or not a stationary equilibrium can explain bargaining behavior in real situations well. This question needs to be investigated empirically. Especially, experimental investigations would be useful. The exploration to this direction is beyond the scope of the present paper. Here, we investigate theoretically what features in bargainers' behavior yield a unique outcome of negotiations with incomplete information. The outcome is expected to be served as a reference point for analyses of mechanism bargaining with incomplete information based on a non-cooperative game theory.

3 Existence

In this section, we prove that there exists a stationary equilibrium of the bargaining game Γ^δ for every δ. In the equilibrium, the ex post Nash bargaining solution is agreed in the initial round with probability one in the limit as δ goes to one.

The first lemma shows the existence of a solution of the well-known equilibrium condition for a subgame perfect equilibrium in the bargaining game Γ^δ in the case of complete information.

Lemma 3.1. For every $i, j = 1, 2$ $(i \neq j)$ and $t \in T$, there exist some real numbers $v_i(t)$ and $w_i(t)$ in R_+ which satisfy the following:

(i) $H^t(w_i(t), \delta v_j(t)) = 0$,

(ii) $v_i(t) = p_i w_i(t) + (1 - p_i)\delta v_i(t)$,

where p_i is the probability that player i is selected as a proposer, and H^t is the Pareto frontier function of $U(t)$.

Proof. For every $t \in T$, let h_i^t be the implicit function of H^t defined by $H^t(h_i^t(u_j), u_j) = 0$. Assumption 2.1 guarantees that h_i^t is well-defined and continuous on the projection of $U(t)_+ = U(t) \cap R_+$ to the j-axis. Define $g_i(u_i, u_j) = p_i h_i^t(\delta u_j) + (1 - p_i)\delta u_i$ for $i = 1, 2$. Then, $g(u) = (g_1(u), g_2(u))$ is a continuous function from the convex set $U(t)_+$ to itself. Since $U(t)_+$ is also a compact set, there exists a fixed point $v^*(t) = (v_1^*(t), v_2^*(t))$ of g by Brouwer's fixed point theorem. Define $w_i^*(t) = h_i^t(\delta v_j^*(t))$ for $i, j = 1, 2$ $(i \neq j)$. Clearly, $v^*(t)$ and $w^*(t)$ satisfy (i) and (ii) in the lemma. Q.E.D.

Two properties (i) and (ii) in the lemma are interpreted as follows. For every $i = 1, 2$ and $t \in T$, $v_i(t)$ means the expected payoff of player i for a subgame perfect equilibrium of Γ^δ in the case of complete information, and

$w_i(t)$ does his demand payoff.[8] Property (i) means that given a type profile t, player i proposes a payoff allocation which maximizes his payoff subject to the constraint that player j receives at least his continuation payoff $\delta v_j(t)$ in case of rejection. The constraint is binding in equilibrium. The expected payoff of player i satisfies property (ii) under the random proposer rule.

Theorem 3.1. For every $\delta < 1$, there exists a stationary equilibrium (σ, μ) of Γ^δ such that every player $i = 1, 2$ proposes a mechanism x_i^δ, independent of his type, and that x_i^δ is accepted by every type of player $j(\neq i)$. As δ goes to one, the sequence $\{x_i^\delta\}$ of mechanisms for every $i = 1, 2$ converges to the ex post Nash bargaining solution x^{NB} with weights $p = (p_1, p_2)$, where p is the probability distribution for proposer selection.

Proof. For every $i, j = 1, 2$ $(i \neq j)$ and $t \in T$, choose $x_i^\delta(t) \in A$ such that $w_i(t) = u_i(x_i^\delta(t), t)$ and $\delta v_j(t) = u_j(x_i^\delta(t), t)$ where $v(t)$ and $w(t)$ satisfy (i) and (ii) in Lemma 3.1. The existence of such $x_i^\delta(t)$ is guaranteed by Lemma 3.1. Let x_i^δ be the mechanism that assigns $x_i^\delta(t)$ to every $t \in T$. Using each x_i^δ, we construct a strategy σ and a belief μ as follows. For every $i, j = 1, 2$ $(i \neq j)$,

(E1) i proposes x_i^δ and j accepts it, independent of their types and a history of play.

(E2) Every player has a belief which satisfies:

- **(a)** when every type t_j of j responds to x_i^δ in the first round, he has the posterior belief $\pi(\cdot|t_j)$ about type t_i of i,
- **(b)** when type t_j responds to any mechanism $y \neq x_i^\delta$ in the first round, he has the posterior belief $\pi(\cdot|T_i^+, t_j)$ over the set $T_i^+ = \{t_i \in T_i | u_i(y(t), t) > w_i(t)\}$ where T_i^+ is a non-empty set, and

[8] In the case of complete information, it is well-known that a subgame perfect equilibrium of the two-person bargaining game Γ^δ is stationary.

(c) after the first round, the same rules as (a) and (b) are applied to a responder's belief where his prior belief is possibly updated according to a game play in previous rounds.

(E3) Every type t_j of j responds optimally to a proposal by player i under the belief (E2) and the strategy (E1) with the tie-breaking rule that he accepts it when he is indifferent to a response.

We show that (σ, μ) is a desired equilibrium of Γ^δ in the theorem. First, (E1) implies that σ is a stationary strategy profile, and that every player i has the conditional expected payoff $v_i(t)$ given every $t \in T$ by property (ii) in Lemma 3.1.

Second, responder j's belief (E2a) in the first round is consistent with the strategy σ on equilibrium play since proposer i proposes the same mechanism x_i, independent of his type. Any belief of j is consistent with σ off equilibrium play in the first round. Specifically, j's belief (E2b) is consistent with σ. After the first round, the two players' prior beliefs are possibly updated according to a history of game play. Since the consistency of the updating rules (E2a) and (E2b) holds for any prior belief of the responder, responder's beliefs (E2c) are also consistent with σ.

Third, we show that (E1) prescribes an optimal proposal for every type t_i of every player i. Suppose that type i deviates from (E1) and proposes a mechanism y. Without loss of generality, we can assume that there exists some $t_j^- \in T_j$ such that $u_i(y(t_i, t_j^-), (t_i, t_j^-)) > w_i(t_i, t_j^-)$. Otherwise, type t_i never becomes better off by proposing y, no matter how player j responds to y. By (E2b), type t_j^- believes that the true type of player i must be in the set $T_i^+ = \{t_i' \in T_i | u_i(y(t_i', t_j^-), (t_i', t_j^-)) > w_i(t_i', t_j^-)\}$. Since the payoff vector $(w_i(t_i', t_j^-), \delta v_j(t_i', t_j^-))$ is Pareto efficient in $U(t_i', t_j^-)$ by (i) in Lemma 3.1, it holds that $u_j(y(t_i', t_j^-), (t_i', t_j^-)) < \delta v_j(t_i', t_j^-)$ for every $t_i' \in T_i^+$. Thus, type t_j^- optimally rejects y. The arguments so far show that type t_i never obtains a payoff higher than $w_i(t)$ for any possible type t_j by proposing y. Thus, it is optimal for t_i to propose x_i. By (E3), (σ, μ) prescribes an optimal response for every player. Since the arguments above do not depend on an initial belief of proposer i, it can be applied not only to the first round but also to other rounds in which the proposer's belief may be updated by a history of play.

Finally, we can see from Lemma 3.1 that the allocation assigned by the constructed equilibrium (σ, μ) to every type profile $t \in T$ satisfies the equilibrium condition in the case of complete information. Since the convergence to the ex post Nash bargaining solution in the last part of the theorem can be proved in the standard manner, its proof is given in Appendix. Q.E.D.

It is well-known that there is a large freedom of players' belief when they observe unexpected actions in a sequential equilibrium. In fact, any arbitrary belief of a responder off equilibrium play can be consistent with the proposer's equilibrium strategy in the sense of Kreps and Wilson (1982) in the bargaining game Γ. In the proof of Theorem 3.1, we choose the following belief of a responder. When he receives an unexpected proposal, he believes that given his type, a true type of a proposer should be among those who are better off by doing so, if it is accepted, than in the equilibrium proposal. Since the equilibrium proposal is Pareto efficient for each type profile, the responder will be worse-off than in the equilibrium proposal if he accepts such an unexpected proposal. Thus, all non-equilibrium proposals are rejected under the selected belief of the responder if the proposer attempts to obtain a payoff higher than in the equilibrium proposal.

The next example illustrates the result of Theorem 3.1.

Example 3.1. Consider a two-person bargaining game in which two players have two types, $T_1 = \{t_1, t_1'\}$ for player 1 and $T_2 = \{t_2, t_2'\}$ for player 2. The prior belief of players is given by the uniform distribution on $T = T_1 \times T_2$. The feasible set $U(t)$ for each type profile $t \in T$ is given by

$$\begin{aligned} U(t_1, t_2) = U(t_1', t_2') &= \{(x_1, x_2) \in R^2_+ | \ 2x_1 + x_2 \leq 1\} \\ U(t_1', t_2) = U(t_1, t_2') &= \{(x_1, x_2) \in R^2_+ | \ x_1 + 2x_2 \leq 1\}. \end{aligned}$$

The four possible feasible sets are illustrated in Figure 3.1. In Figure 3.1, U_{12} denotes the feasible set $U(t_1, t_2)$ where player 1 is of type t_1 and player 2 is of type t_2. Other notations of feasible sets can be interpreted similarly.

A story behind the feasible sets is as follows (Harsanyi 1968). Two players negotiate for a division of a fixed amount of money. One of them may have to pay half of his gross payoffs to the third party, depending on their type profile. When the type profile t is either (t_1, t_2) or (t_1', t_2'), it is player 1 who pays. When the type profile t is either (t_1', t_2) or (t_1, t_2'), it is player 2 who pays. No type of a player knows in advance who pays. The disagreement payoffs are $(0, 0)$, independent of players' types. The ex post (symmetric) Nash solution x^{NB} of this two-person Bayesian bargaining problem is given by $x^{NB}(t_1, t_2) = x^{NB}(t_1', t_2') = (\frac{1}{4}, \frac{1}{2})$ and $x^{NB}(t_1', t_2) = x^{NB}(t_1, t_2') = (\frac{1}{2}, \frac{1}{4})$.

According to Theorem 3.1, for a sufficiently large δ, the ex post Nash solution x^{NB} can be asymptotically attained by a sequential equilibrium of the bargaining game Γ^δ where two players are selected as a proposer with equal probability. In the equilibrium, every type of player 1 proposes the mechanism x_1^δ satisfying $x_1^\delta(t_1, t_2) = x_1^\delta(t_1', t_2') = (\frac{2-\delta}{4}, \frac{\delta}{2})$ and $x_1^\delta(t_1', t_2) = x_1^\delta(t_1, t_2') = (\frac{2-\delta}{2}, \frac{\delta}{4})$. Every type of player 2 proposes the mechanism x_2^δ where $x_2^\delta(t_1, t_2) = x_2^\delta(t_1', t_2') = (\frac{\delta}{4}, \frac{2-\delta}{2})$ and $x_2^\delta(t_1', t_2) = x_2^\delta(t_1, t_2') = (\frac{\delta}{2}, \frac{2-\delta}{4})$. When every player is offered a mechanism y, he believes that a true type of the proposer should be among those (if any) who are better off in y than in the equilibrium mechanism, given his type. Owing to the linearity of the Pareto frontier in each feasible set, the expected equilibrium payoff of every player given a type profile t coincides with the Nash bargaining solution $x^{NB}(t)$.

By definition, the ex post Nash bargaining solution is ex post efficient,

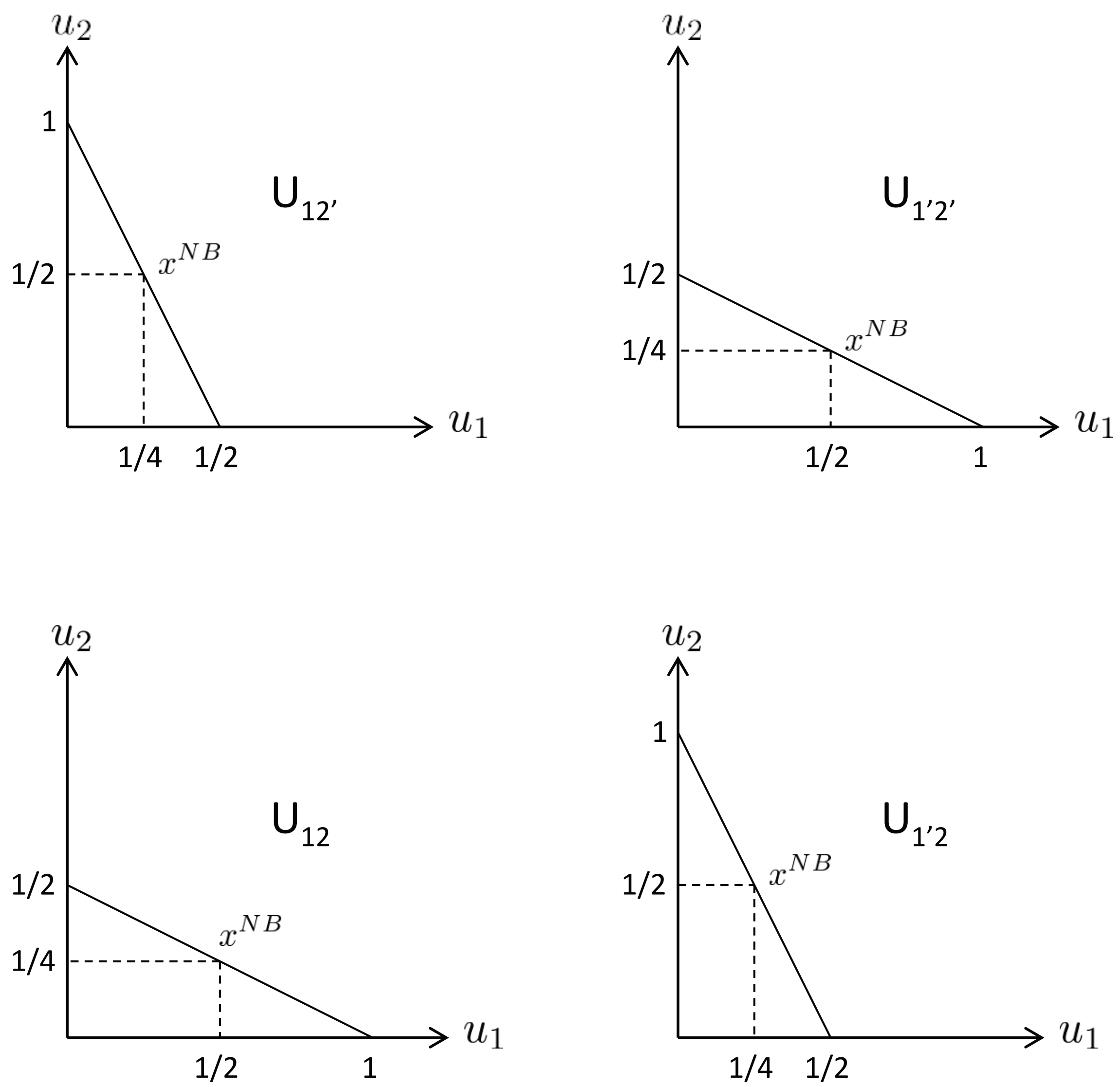

Figure 3.1 Bayesian bargaining problem

that is, it is Pareto efficient given every type profile of players. It, however, is not interim efficient in terms of conditional expected payoffs given every player's type. For example, consider a mechanism y in Example 3.1 such that $y(t_1, t_2) = y(t'_1, t'_2) = (1, 0)$ and $y(t'_1, t_2) = y(t_1, t'_2) = (0, 1)$. The mechanism y assigns the whole payoff to an efficient player who does not need to pay half of his dividend to the third party. The conditional expected payoff of every type of every player is $\frac{1}{2}$ for y, and is $\frac{3}{8}$ for the ex post Nash bargaining solution x^{NB}. This means that y interim-payoff dominates x^{NB}. Then, a natural question arises: why do not two players agree to y? The answer is as follows. If any type, say t_1, of player 1 proposes y, then type t_2 of player 2 believes that a true type of player 1 must be type t_1, knowing that only type t_1 is better off in y than in x^{NB}. Under this updated belief, type t_2 of player 2 optimally rejects y. By the same reason, type t'_2 of player 2 optimally rejects y, too. It can be easily shown that the mechanism y is interim-efficient. So, any cooperative solution assuming interim-efficiency and interim symmetry selects y in the example. The non-cooperative analysis in this section does not support this selection.

In the following sections, we consider under what conditions a stationary equilibrium of the bargaining game Γ can uniquely implement the ex post Nash bargaining solution in the limit as the discount factor for future payoffs goes to one. To answer this question, we examine the three properties of the stationary equilibrium constructed in the proof of Theorem 3.1:

1. (no delay of agreements) an agreement is immediately made with probability one,

2. (inscrutability) every type of a proposer proposes the same mechanism on the equilibrium play, and

3. (information revealing) a proposer may update his prior belief based on revealed information when he receives an unexpected proposal (off the equilibrium play).

4 No-delay Agreements

The timing of agreements has been a major topic in the literature of sequential bargaining. In the case of complete information, Rubinstein (1982) shows that an agreement is reached in the initial round, provided that future payoffs are discounted. On the other hand, a large volume of literature on sequential bargaining with incomplete information shows that delay of agreements may happen in equilibrium. Rubinstein (1985) further considers a sequential bargaining game with one-sided incomplete information where one player has two types, strong and weak, about time preference and the opponent does not know his type. Rubinstein shows that delay may occur with a positive probability. Delay is caused by a conflict between different types of the informed player. A weak type may want to pretend to be a strong type. To gain an advantage over the weak type, a strong type may want to reveal his type by making an unacceptable offer. Equilibrium delay has been investigated in many other bargaining models with incomplete information (Fudenberg and Tirole 1983; Chatterjee and Samuelson 1987; Grossman and Perry 1986b, for example). In fact, most previous studies are motivated to explain delay in reaching agreements among rational agents.

An opposite approach is taken here. We are concerned with under what conditions an agreement can be made immediately with probability one in negotiations with incomplete information. To consider this problem, it is important to notice a difference between bargaining over allocations studied in the literature and that over mechanisms in this paper. In the mechanism bargaining, players negotiate over allocations contingent on every type profile of players. In other words, players can negotiate on an allocation "type by type," and thus the competition among different types of the same player does not have a direct effect on an agreement. It, however, has an indirect effect on an agreement since the opponent is uncertain about a true type of the player.

Due to an informational linkage among different types, it is not simple to answer the question whether an agreement is made immediately in mechanism bargaining. A proposal may affect a belief of a responder about a type of the opponent, and the responder may have a belief such that rejection is his optimal response. In a simple case of ultimatum bargaining (corresponding to the case of $\delta = 0$), we show that a proposal may be rejected with a positive probability in a sequential equilibrium. Note that an agreement is reached in the case of complete information as long as there is a mutually beneficial allocation for players.

Example 4.1. Consider a two-person bargaining game in which two players have two types, $T_1 = \{t_1, t'_1\}$ for player 1 and $T_2 = \{t_2, t'_2\}$ for player 2. The prior belief of players is the uniform distribution on $T_1 \times T_2$. Player 1 is called player $1(t_1)$ if he is of type t_1. Similar notations are used for other types of players. Consider two mechanisms x and y in Table 4.1. For simplicity, it is assumed that other mechanisms are not feasible. Note that the two mechanisms x and y assign the same payoffs for both players when player 2 is of type t'_2. We consider the ultimatum protocol whereby player 1 proposes either x or y and player 2 responds to this. If player 2 rejects the proposal, then the game ends with the disagreement payoffs $(0, 0)$.

We construct a sequential equilibrium where a proposal may be rejected. Player 1 proposes x, independent of his type. Player $2(t_2)$ rejects x and player $2(t'_2)$ accepts it. If player 1 proposes y, then player 2 responds in the opposite way, that is, player $2(t_2)$ accepts y and player $2(t'_2)$ rejects it. Player 2's beliefs are given as follows. Since proposal x gives player 2 no additional information about player 1, all types of player 2 receiving x believe that player 1 is of either type t_1 or type t'_1 with equal probability. Given proposal y, player $2(t_2)$ has an arbitrary belief, and player $2(t'_2)$ believes that player 1 must be of type t'_1. Those beliefs of player 2 off equilibrium play are consistent with player 1's strategy in the sense of Kreps and Wilson (1982). It can be easily seen that player 2's strategy prescribes his optimal responses for all his types under the belief specified above. Given player 2's responses, player $1(t_1)$ receives expected payoff 3 for x and 2 for y. Similarly, player $1(t_1')$ receives expected payoff 3/2 for x and 1 for y. Thus, x is the optimal proposal for every type of player 1.

	t_2	t_2'
t_1	$1, -2$	$6, 6$
t_1'	$1, 1$	$3, -1$

x

	t_2	t_2'
t_1	$4, 4$	$6, 6$
t_1'	$2, 2$	$3, -1$

y

Table 4.1 Two mechanisms in ultimatum bargaining

In equilibrium, player 1's proposal x is rejected by player $2(t_2)$. One may wonder why any type of player 1 does not propose y which makes all types of all players weakly better off than in x. The reason is that proposing y affects player $2(t_2')$'s belief and, as a result, he rejects y. Thus, each type of player 1 is worse off by proposing y than by proposing x.

There exists another sequential equilibrium where an agreement is made with probability one. In equilibrium, player 1 proposes mechanism y, independent of his type. All types of player 2 accept it. Player $2(t_2)$ rejects x, and player $2(t_2')$ accepts it. The beliefs of player 2 are as follows. Given proposal y, all types of player 2 believe that player 1 may be of either type t_1 or type t_2 with equal probability. Given proposal x, they believe that player 1 is of type t_1. It is easy to see that the strategy is a sequential equilibrium under these beliefs.

Two mechanisms x and y in Table 4.1 are identical from the viewpoint of player $2(t_2')$: he knows that the two mechanisms yield the same outcome. In the second equilibrium with an agreement, player $2(t_2')$ responds to x and y in the same manner, that is, acceptance. On the contrary, he responds to them differently in the first equilibrium where an agreement may not be reached with a positive probability. In what follows, we prove that there exists no delay of agreements in a stationary equilibrium of the bargaining game Γ if every type of a player responds identically to two proposals which are identical, given his type.

For a stationary equilibrium (σ, μ) of Γ, let $M(\sigma)$ be the set of all mechanisms proposed on equilibrium plays of σ. We consider the following property of responders' behavior.

Definition 4.1. A stationary equilibrium (σ, μ) of Γ is said to satisfy *independence of irrelevant types* (IIT) if, for every $i = 1, 2$, $t_i \in T_i$, $x \in M(\sigma)$, and $y \in M$,

$$x = y \text{ on } T(t_i) \text{ implies } \sigma_i(t_i, x) = \sigma_i(t_i, y),$$

where $\sigma_i(t_i, x)$ and $\sigma_i(t_i, y)$ are the responses of player i to x and y, respectively, prescribed by σ_i when his type is t_i.

The IIT condition means that every type of a player responds to an equilibrium proposal and a non-equilibrium proposal in the same way if they prescribe the same outcomes, given his type, in every contingency for the other player's type. In other words, every type of a player makes the same responses to two mechanisms if *he knows* that they are identical. Every player type's response to a proposal is independent of the allocations it assigns to his other (irrelevant) types.

Proposition 4.1. If a stationary equilibrium (σ, μ) of Γ satisfies IIT, then every player's proposal is accepted in the initial round with probability one.

Proof. Given a type profile $t \in T$ for the players, let $v(t) = (v_1(t), v_2(t))$ be their conditional expected discounted payoffs for σ evaluated at the start of each round before the random selection of a proposer. Since σ is stationary, $v(t)$ is independent of past actions. It holds that $v(t) \in U(t)$ since $U(t)$ is a closed and convex set by Assumption 2.1.(1). Since the disagreement payoff $u(d^*, t) = (0, 0)$ is an interior point of $U(t)$ by Assumption 2.1.(2), it holds that $\delta v(t)$ is also an interior point of the convex set $U(t)$.

By way of contradiction, suppose that there exists some player i, say $i = 1$, whose equilibrium proposal $x \in M(\sigma)$ is rejected with positive probability in the initial round in σ when his type is some $t_1^* \in T_1$. Then, the type set T_2 of player 2 is partitioned into two subsets, T_2^a and T_2^r, such that x is accepted on $\{t_1^*\} \times T_2^a$ and rejected on $\{t_1^*\} \times T_2^r$ in σ. T_2^r is non-empty by supposition. For type t_1^* of player 1, his equilibrium proposal x is rejected by each type $t_2 \in T_2^r$ of player 2, and the game goes to the next round. Thereafter, the continuation payoffs for the two players with type profile $t \in T_1 \times T_2^r$ are given by $\delta v(t)$ since σ is stationary. Since $\delta v(t)$ is an interior point of $U(t)$ for all $t \in T$, there exists a mechanism $y \in M$ such that

(i) $u_j(y(t), t) > \delta v_j(t)$ for every $j = 1, 2$ and every $t \in T_1 \times T_2^r$,

(ii) $y(t) = x(t)$ for every $t \in T_1 \times T_2^a$.

Suppose that player 1 employs strategy $\sigma_1'(\neq \sigma_1)$ to propose y when he is of type t_1^*. For every $t_2 \in T_2^r$, it holds by (i) that for every $t = (t_1, t_2) \in T(t_2)$,

$$u_2(y(t), t) > \delta v_2(t).$$

Thus, every type $t_2 \in T_2^r$ of player 2 optimally accepts y, independent of his belief. For every $t_2 \in T_2^a$, it holds by (ii) that for every $s_1 \in T_1$,

$$y(s_1, t_2) = x(s_1, t_2).$$

By IIT, it holds that $\sigma_2(t_2, x) = \sigma_2(t_2, y)$. Since $t_2 \in T_2^a$, $\sigma_2(t_2, x) =$ accept. Thus, every type $t_2 \in T_2^a$ of player 2 accepts y in σ.

It has been shown that all types of player 2 accept proposal y by type t_1^* of player 1. Thus, the conditional expected payoff for player 1 given t_1^* for

(σ_1', σ_2) satisfies

$$\begin{aligned} Eu_1(\sigma_1', \sigma_2 | t_1^*) &= \sum_{t_2 \in T_2^r} \pi(t_2|t_1^*) u_1(y(t), t) + \sum_{t_2 \in T_2^a} \pi(t_2|t_1^*) u_1(y(t), t) \\ &> \sum_{t_2 \in T_2^r} \pi(t_2|t_1^*) \delta v_1(t) + \sum_{t_2 \in T_2^a} \pi(t_2|t_1^*) u_1(x(t), t) \qquad (3) \\ &= Eu_1(\sigma_1, \sigma_2 | t_1^*) \end{aligned}$$

where $t = (t_1^*, t_2)$. This contradicts the fact that σ is a sequential equilibrium. Q.E.D.

The roles of stationarity and IIT in the proposition can be explained as follows. If negotiations fail between two players with a type profile $t \in T$, then each player $i (= 1, 2)$ expects to receive the continuation payoff $\delta v_i(t)$ where $v_i(t)$ is player i's conditional expected payoff given t, evaluated at the beginning of each round. Since an equilibrium is stationary, $v_i(t)$ is independent of a history of game play. Suppose that some type t_i^* of player i, say $i = 1$, makes an unacceptable proposal x in the initial round. Then, player 2 are divided into two types, those who accept x (T_2^a in the proof) and those who reject (T_2^r in the proof). Since the continuation payoff vector $\delta v(t)$ is in the interior of the feasible set $U(t)$ for all $t \in T$, type t_1^* of player 1 can construct and propose a new mechanism y such that (i) players 1 and 2 are strictly better off in y than in $\delta v(t)$ for any type profile t in $T_1 \times T_2^r$, and (ii) x and y are identical on $T_1 \times T_2^a$. Property (i) implies that all rejection types T_2^r of player 2 accept y, regardless of their beliefs about player 1's type. Note that rejection types in T_2^r do not know whether or not non-equilibrium proposal y is made by type t_1^* of player 1. Property (ii) means that all acceptance types in T_2^a know that x and y prescribe the same outcomes. Thus, IIT implies that they respond to x and y in the same manner, that is, they accept y. Since all types of player 2 accept y, type t_1^* of player 1 is better off if he proposes the non-equilibrium mechanism y. This is a contradiction.

To conclude the section, let us discuss the relevance of the IIT condition. The condition assumes that every type of a responder responds to a non-equilibrium proposal in the same way as to an equilibrium one if both proposals are identical, given his type. IIT trivially holds if the responder has the same beliefs about a true type of the opponent when he receives either of the two proposals. We, however, remark that IIT does not necessarily assume this. Specifically, consider again the sequential equilibrium constructed in Theorem 3.1 that implements asymptotically the ex post Nash bargaining solution. In equilibrium, every type of a responder is offered exactly his continuation payoff in every contingency of players' types. This implies that the responder optimally accepts a non-equilibrium proposal in the IIT condition under any belief about the opponent. Thus, IIT is satisfied without any restriction on a responder's belief in the case of the ex post Nash bargaining solution. In a general case, IIT may restrict a responder's belief off the equilibrium so that the response assumed by it can be optimal to him. For example, the second equilibrium in Example 4.1 satisfies IIT if type t_2' of player 2 believes that player 1 may be of type t_1 with at least probability 1/7, being proposed mechanism x. Under the prior belief, he anticipates so with probability 1/2. Note that IIT does not violate the notion of a sequential equilibrium since it allows an arbitrary belief of the responder off the equilibrium play in Γ.

5 Characterization

In this section, we first show that there is no loss of generality if we restrict our analysis to a pooling equilibrium where all types of proposer choose the same mechanisms. In such an equilibrium, the choice of a mechanism does not reveal any private information of the proposer. Myerson (1983) calls this result the *principle of inscrutability* and justifies it in his ultimatum bargaining model of an informed principal. The following lemma shows that the principle also holds true in the sequential bargaining game Γ.

Lemma 5.1. For any stationary equilibrium (σ, μ) of Γ satisfying IIT, there exists some stationary equilibrium (σ', μ') of Γ that satisfies IIT and the following properties:

(i) (σ, μ) and (σ', μ') are *outcome-equivalent*; that is, both equilibria generate the same outcomes for every type profile $t \in T$.

(ii) In (σ', μ'), all types of every player $i = 1, 2$ propose the same mechanism $x_i^* \in M$. The other player accepts it, independent of his type.

The proof of the lemma is given in Appendix. The result was first proved by Myerson (1983) in a problem of mechanism design by an informed principal. Although we assume IIT for the sake of our analysis, the lemma is a general principle which holds without IIT in a mechanism bargaining game (see Okada 2012). The basic idea of Myerson can be applied to a general situation.

A key observation to prove Lemma 5.1 is that any equilibrium generates a single mechanism that assigns the same outcome as in equilibrium to every type profile of players. When different types of the proposer propose different mechanisms, such a single mechanism can be defined by "combining" different mechanisms over the proposer's type set. Then, we can construct a new equilibrium in which all types of the proposer propose this outcome-equivalent mechanism. All types of the responder accept it under the posterior beliefs, knowing their own types. Off the equilibrium play, the new equilibrium coincides with the original one. If any private information regarding the proposer is revealed in the original equilibrium, then it is optimal for the responder to accept the proposal, given his type and each of the revealed information. Since the acceptance is optimal for the responder given *every* revealed information in the original equilibrium, it is also optimal for him to accept the proposal in the new equilibrium where no information is revealed. Given the responder's acceptance, each type of the proposer is indifferent to which he proposes, the original mechanism or the constructed one.

In addition to the inscrutability principle, we need a refinement of a sequential equilibrium for our characterization result. It is well-known that many sequential bargaining games have a large set of sequential equilibrium outcomes, caused by a freedom of players' beliefs off the equilibrium play. Specifically, a responder's belief about a proposer can be arbitrary in a sequential equilibrium when he is offered an unexpected proposal. Some of responders' beliefs, however, are unreasonable in the situation that the proposer has an incentive to screen himself. To eliminate unreasonable beliefs off equilibrium play, we introduce a self-selection condition which has been considered in the literature of refinements of sequential equilibrium (Grossman and Perry 1986a and 1986b and Rubinstein 1985 among others).

To illustrate the idea of our refinement, let us consider again the sequential equilibrium in Example 3.1 that implements the ex post Nash bargaining solution. In equilibrium, every type of player 1 proposes the mechanism x_1^δ satisfying $x_1^\delta(t_1,t_2) = x_1^\delta(t_1',t_2') = (\frac{2-\delta}{4}, \frac{\delta}{2})$ and $x_1^\delta(t_1',t_2) = x_1^\delta(t_1,t_2') = (\frac{2-\delta}{2}, \frac{\delta}{4})$. Suppose that type t_1 of player 1 proposes a non-equilibrium mechanism y such that $y(t_1,t_2) = y(t_1',t_2') = (1,0)$ and $y(t_1',t_2) = y(t_1,t_2') = (0,1)$. Although type t_2 of player 2 does not know a true type of player 1, either t_1 or t_1', he knows that only type t_1 is better off in y, if it is accepted, than in the equilibrium proposal x_1^δ. With this knowledge, type t_2 of player 2 infers credibly that a true type of player 1 must be t_1, not t_1'. Our refinement requires that type t_2 of player 2 should have such a belief, given the proposal y.

We now formalize the self selection property of a sequential equilibrium.

Definition 5.1. Let (σ,μ) be a stationary equilibrium of Γ satisfying IIT in which every player $i=1,2$ proposes a mechanism x_i (independent of his type). An equilibrium (σ,μ) is said to satisfy *self-selection* if, when every type $t_j \in T_j$ of responder $j(\neq i)$ receives a proposal y_i from player i satisfying that the set

$$T_i^+ = \{t_i \in T_i | u_i(y_i(t),t) > u_i(x_i(t),t) \ \text{ for } t=(t_i,t_j)\}$$

is non-empty, the belief system μ assigns to type t_j of responder j a posterior belief of which support is equal to T_i^+.[9] If T_i^+ is an empty set, then no restriction on the belief system is imposed.

The property of self-selection can be explained as follows. Suppose that a responder receives an unexpected proposal off equilibrium play. It assumes that the responder believes that a true type of the proposer should be among those (T_i^+) who are better off by the proposal, if it is accepted, than in the equilibrium proposal, given his type. In other words, the proposer credibly reveals his type in T_i^+ by making a non-equilibrium proposal where all types of T_i^+ and only themselves have incentives to doing so. Note that our self-selection property is weak in the sense that it does not restrict a responder's belief to his posterior belief $\pi(\cdot|T_i^+, t_j)$ given (T_i^+, t_j), allowing an arbitrary belief with support T_i^+.

The self-selection belief gives us the following refinement test of a sequential equilibrium. Suppose that some type of a proposer deviates from the equilibrium, and that he makes a non-equilibrium proposal. If all types of the responder accept it under their self-selection beliefs and thus the deviating type of the proposer becomes better off, then the sequential equilibrium in question is considered to be destabilized by the deviation. We eliminate such an unreasonable equilibrium.

A refinement of a sequential equilibrium based on the idea of self-selection is first proposed by Grossman and Perry (1986a) in the context of the two-player signaling games. They name an equilibrium satisfying self-selection a perfect sequential equilibrium. Grossman and Perry show that the self-selection property is stronger than the criterion of Cho and Kreps (1987) for signaling games. The self-selection is also related to "neologism-proofness" of Farrell (1993) for cheap-talk games. While both refinements of a sequential equilib-

[9] The support of a probability distribution F on the finite set T_i is the set of all elements $t_i \in T_i$ with $F(t_i) > 0$.

rium impose some restrictions of receivers' beliefs, a difference between them is that the criterion of Farrell essentially allows a sender to choose an updating rule which is in his best interest, assuming the mutual understanding of meaning of language. See Grossman and Perry (1986a) on this point. Rubinstein (1985) and Grossman and Perry (1986b) show that the self-selection refinement is so powerful that it selects a unique sequential equilibrium in two-person alternating-offers bargaining games with one-sided incomplete information. In Okada (2012), we present a refinement of Wilson's (1978) coarse core, called the signaling core, of an n-person coalitional game with incomplete information based on a criterion of self-selection.

The next is a key lemma for our characterization result.

Lemma 5.2. Suppose that (σ, μ) is a stationary equilibrium of Γ satisfying IIT and self-selection. For every $i = 1, 2$, let x_i be the equilibrium mechanism proposed by every type of player i, and let $v_i(t)$ be the conditional expected payoff $Eu_i(\sigma|t)$ of player i for σ at the beginning of each round given t. Then, for every $i, j = 1, 2$ $(i \neq j)$ and every $t \in T$, the following properties hold:

(i) $u_j(x_i(t), t) = \delta v_j(t)$, and

(ii) $u(t) = (u_i(x_i(t), t), u_j(x_i(t), t))$ is Pareto efficient in $U(t)$.

Proof. (i) It follows from the inscrutability principle (Lemma 5.1) that player i proposes x_i in (σ, μ), independent of his type. Thus, responder j never receives additional information from x_i, and so he does not update the prior belief π. Since every type t_j of player j accepts x_i by Proposition 4.1, it must hold that

$$\sum_{t_i \in T_i} \pi(t) u_j(x_i(t), t) \geq \sum_{t_i \in T_i} \pi(t) \delta v_j(t). \tag{4}$$

It suffices us to show that $u_j(x_i(t), t) \leq \delta v_j(t)$ for every $t = (t_i, t_j)$. If this is the case, then we have $\sum_{t_i \in T_i} \pi(t) u_j(x_i(t), t) \leq \sum_{t_i \in T_i} \pi(t) \delta v_j(t)$. Since the opposite inequality also holds true by (4), we can conclude that $u_j(x_i(t), t) =$

$\delta v_j(t)$ for every t. That is, (i) holds.

By way of contradiction, suppose that $u_j(x_i(s), s) > \delta v_j(s)$ for some $s \in T$. If the payoff vector $u(s) = (u_i(x_i(s), s), u_j(x_i(s), s))$ is on the Pareto frontier of $U(s)$, then Assumption 2.1.(1) guarantees that there exists an action $a \in A$ such that $u_i(a, s) > u_i(x_i(s), s)$ and $u_j(x_i(s), s) > u_j(a, s) > \delta v_j(s)$ by making a slight "payoff transfer" between i and j at $u(s)$ along the Pareto frontier of $U(s)$. If $u(s)$ is not on the Pareto frontier of $U(s)$, then it is clear that there exists an action $a \in A$ such that $u_i(a, s) > u_i(x_i(s), s)$ and $u_j(a, s) > u_j(x_i(s), s) > \delta v_j(s)$. Consider the mechanism y_i that assigns the action a to s and coincides with x_i for all other type profiles. Then y_i satisfies

$$u_i(y_i(s), s) > u_i(x_i(s), s) \tag{5}$$

$$u_i(y_i(t), t) = u_i(x_i(t), t) \quad \text{for every } t \neq s, \tag{6}$$

$$u_j(y_i(s), s) > \delta v_j(s). \tag{7}$$

Since (σ, μ) satisfies self-selection, it follows from (5) and (6) that type s_j of player j believes that the true type of player i must be s_i, if type s_i of player i proposes y_i. By (7), type s_j optimally accepts y_i. For all other types of j, y_i prescribes the same actions as x_i. Thus, IIT requires that they should respond to y_i in the same way as to x_i. That is, they accept the proposal. Since all types of j accept y_i, (5) implies that type s_i of player i is better off by proposing y_i in (σ, μ) than x_i. This is a contradiction that (σ, μ) is a sequential equilibrium of Γ.

(ii) By way of contradiction, suppose that $u(s)$ is not Pareto efficient in $U(s)$ for some $s \in T$. Then there exists some $u' = (u'_i, u'_j) \in U(s)$ such that $u'_i > u_i(x_i(s), s)$ and $u'_j > u_j(x_i(s), s) = \delta v_j(s)$. The last equality comes from (i). Similarly to the proof of (i), consider the mechanism y_i that assigns the action yielding payoffs u' to s and coincides with x_i for all other type profiles. Then, y_i satisfies $u_j(y_i(s), s) = u'_j > \delta v_j(s)$ and (5) and (6). By the same arguments as in (i), if type s_i of player i proposes y_i, then all types of player j

accept it, and thus type s_i is better off than in (σ, μ). This is a contradiction that (σ, μ) is a sequential equilibrium of Γ. Q.E.D.

The lemma shows that a stationary equilibrium with IIT and self-selection in Γ necessarily satisfies the equilibrium condition in the case of complete information. Specifically, for every type profile t, proposer i offers responder j exactly his continuation payoff $\delta v_j(t)$, being equal to the discounted value of his conditional expected payoff $v_i(t)$ given t. The logic for this result can be explained as follows. By the no-delay result (Proposition 4.1) and the inscrutability principle (Lemma 5.1), it holds that every type of proposer i proposes the same mechanism and every type of responder j accepts it. This implies that responder j's conditional expected payoff for the equilibrium mechanism given his every type t_j is greater than or equal to the conditional expected value of $\delta v_j(t)$ given t_j. A responder receives no additional information about the type of a proposer. Then, there are two possibilities: (a) the equilibrium offer to responder j is equal to his continuation payoff $\delta v_j(t)$ for every type profile t, and (b) the equilibrium offer to j is strictly greater than $\delta v_j(s)$ for some type profile s. Suppose that case (b) happens. Then, by decreasing the offer to j slightly at s, the proposer can construct a new mechanism whereby he is better off than his equilibrium payoff at s and the responder is still better than $\delta v_j(s)$, while the new mechanism coincides with the equilibrium one for all other type profiles. If type s_i of proposer i makes this new proposal, then type s_j of responder j believes that proposer i must be of type s_i, according to the self-selection property. As a result, responder type s_j accepts the new proposal, since he is better off by doing so than $\delta v_j(s)$. Moreover, all other responder types also accept it by IIT since both the new and the equilibrium mechanisms assign the same outcomes to them. Since all possible responder types accept the new proposal, type s_i of proposer i is actually better off by proposing it. This is a contradiction. By the same logic, it can be shown that the equilibrium mechanism assigns a Pareto efficient outcome to every type profile.

The following theorem characterizes a stationary equilibrium satisfying IIT and self-selection in Γ.

Theorem 5.1. Every player $i = 1, 2$ proposes a mechanism x_i, independent of his type, in a stationary equilibrium (σ, μ) of Γ^δ satisfying IIT and self-selection if and only if x_1 and x_2 satisfy the following properties for every $t \in T$: for $j \neq i$,

(i) $w_i(t) = u_i(x_i(t), t), \delta v_i(t) = u_i(x_j(t), t)$

(ii) $v_i(t) = p_i w_i(t) + (1 - p_i)\delta v_i(t)$,

(iii) $H^t(w_i(t), \delta v_j(t)) = 0$,

where p_i is the probability that player i is selected as a proposer, and H^t is the Pareto frontier function of $U(t)$. The equilibrium mechanism x_i proposed by player i converges to the ex post Nash bargaining solution x^{NB} as δ goes to one.

Proof. The "only if" part follows from Lemma 5.2. To prove the "if" part, it suffices to show that the stationary equilibrium (σ, μ) constructed in the proof of Theorem 3.1 satisfies IIT and self-selection. In σ, only x_i $(i = 1, 2)$ are proposed, that is, $M(\sigma) = \{x_1, x_2\}$. If any type t_j of player j is offered a mechanism y satisfying $y = x_i$ on $T(t_j)$, then he receives payoff $\delta v_j(t)$ no matter how he responds, for every possible type t_i. Thus, type t_j is indifferent to whether he should accept or reject y, independent of his belief about type t_i. According to (E3) in the proof of Theorem 3.1, type t_j accepts y by the tie-breaking rule in σ. This means that (σ, μ) satisfies IIT. The belief μ prescribed by (E2b) and (E2c) clearly satisfies self-selection. The convergence result is proved by Theorem 3.1 (see the proof in Appendix). Q.E.D.

We summarize the characterization result. When two players are sufficiently patient, they agree to the ex post Nash bargaining solution in the first round in the bargaining game Γ, regardless of who proposes, if and only if their behavior is described by a stationary sequential equilibrium satisfying IIT and self-selection.

6 Discussion

The first result (Theorem 3.1) shows the existence of a sequential equilibrium in a sequential bargaining game of mechanism selection in which the ex post Nash bargaining solution is immediately agreed, independent of players' types, in the limit as the discount factor (or the continuation probability) goes to one. Equilibrium strategies are stationary. Since the ex post Nash bargaining solution is not interim efficient in general, the result implies that the axiom of interim efficiency assumed in the cooperative solutions with incomplete information introduced by Harsanyi and Selten (1972) and Myerson (1984) is not always supported in a non-cooperative approach to the Bayesian bargaining problem. The sequential equilibrium constructed in the proof involves a responder's punishment (rejection) by his posterior belief based on self-selection when a proposer chooses a non-equilibrium mechanism. If such an unexpected proposal is made, then every type of the responder rationally infers that a true type of the proposer must be one of those who become better off by the proposal, if it is accepted, than in the equilibrium proposal, given his type. Since the ex post Nash bargaining solution is Pareto efficient for every type profile of players, the responder would be worse off against all such types of the proposer, and thus he optimally rejects the non-equilibrium mechanism.

The characterization result (Theorem 5.1) strengthens the implication of the paper. It shows that the ex post Nash bargaining solution is an asymptotically unique outcome of the Bayesian bargaining problem if and only if bargaining behavior of players is described by a stationary sequential equilibrium satisfying IIT and self-selection. To obtain the characterization result, we have first proved the no-delay result of agreements. IIT plays a critical role in the proof. It restricts the behavior of every type of a responder so that his response to a proposal is independent of the allocations it assigns to his all other types. While IIT implicitly imposes some restrictions on the responder's belief off equilibrium play in a general case, it does not so for the ex post Nash bargaining solution. IIT holds true for any belief in the ex post Nash bargaining solution. Given the no-delay result, the refinement of a sequential equilibrium by self-selection enables us to obtain the equilibrium condition in the case of complete information that a responder is exactly offered his continuation payoff for every type profile of players. If there exists any type profile for which the responder receives strictly higher payoff than his continuation payoff, then the proposer makes a new mechanism where his type is revealed to the responder by the self-selection and he is better off than in the equilibrium, while the responder remains to be better off than his continuation payoff. The acceptance of the new mechanism is guaranteed by the construction of it and IIT.

The result of the paper has the following implication to economic analysis of insurance contracts. Insurance benefit is impossible if it is contingent solely on private information possessed by players. Even if one player proposes an insurance contract which makes all players better off than the ex post Nash solution (in terms of conditional expected payoffs given their own types) at the interim stage, then some private information about the proposer may be revealed by the proposal itself, and the responder optimally rejects it under the revealed information. Insurance contract should be designed so that it is contingent on common risks to all players.

Finally, we discuss some extensions of our analysis. The results of the paper can be extended without much difficulty to the case of $n(> 3)$ players if no coalition of players is allowed. Although we use a particular bargaining protocol with random proposers, the results hold for the alternating-offers model. The analysis of the paper is restricted to a stationary equilibrium.

While a stationary equilibrium can be served as a useful reference point for our analysis of mechanism bargaining with incomplete information, it is interesting to analyze a non-stationary equilibrium of the bargaining model.

The assumption of verifiable types is certainly a limitation of our analysis. When players' types are unverifiable, the bargaining model should be expanded so that an agreement of a contract is followed by a communication game (in the case of a direct mechanism) where all players report their types to an arbitrator who implements the contract. The whole process of negotiations and implementation should be analyzed as a non-cooperative game. The analysis of this paper suggests that IIT and the self-selection refinement would be useful to analyze such an extended game, too. In particular, the Bayesian incentive compatibility condition may be modified so that it could take into account the possibility of information revealing in negotiations. The extension of the analysis to the case of unverifiable types will be an interesting work for future research. If such an extension is successfully done, two branches in game theory, non-cooperative bargaining theory and mechanism design theory, will become closer.

7 Conclusion

We have presented a non-cooperative two-person sequential bargaining game with incomplete information in which players negotiate for mechanisms with verifiable types. We have proved that there exists a stationary sequential equilibrium of the bargaining game in which the ex post Nash bargaining solution with no delay is asymptotically implemented with probability one. We have further proved that the ex post Nash bargaining solution is an asymptotically unique outcome of a stationary sequential equilibrium satisfying IIT and self-selection. Information revealing in negotiations prevents the interim efficiency of an agreement. The paper extends the non-cooperative bargaining theory with complete information to a general case of incomplete information.

Appendix

For simplicity of exposition, we prove the last part of Theorem 3.1 in the case that the Pareto frontier function H^t is differentiable. In the bargaining theory with complete information, it is well-known that the convergence holds true in a non-differentiable case, too. Our proof is based on Okada (2010).

Proof of the last part in Theorem 3.1. For every $i = 1, 2$ and $t \in T$, let $v_i^\delta(t)$ and $w_i^\delta(t)$ satisfy (i) and (ii) in Lemma 3.1. Then, it holds for every $t \in T$ that

$$H^t(w_1^\delta(t), \delta v_2^\delta(t)) = 0, \quad \text{and} \quad H^t(\delta v_1^\delta(t), w_2^\delta(t)) = 0, \tag{8}$$

where H^t is the Pareto frontier function of the feasible set $U(t)$. Let $z_1^\delta(t) = (w_1^\delta(t), \delta v_2^\delta(t))$ and $z_2^\delta(t) = (\delta v_1^\delta(t), w_2^\delta(t))$. $z_i^\delta(t)$ is the payoff vector that the mechanism x_i^δ assigns to $t \in T$. Then, from (8) we have

$$H^t(z_1^\delta(t)) - H^t(z_2^\delta(t)) = 0.$$

By Taylor's theorem, there exists some λ, $0 < \lambda < 1$, such that

$$\begin{aligned}
&[w_1^\delta(t) - \delta v_1^\delta(t)] \cdot \frac{\partial H^t}{\partial x_1}(\lambda z_1^\delta(t) + (1-\lambda) z_2^\delta(t)) + \\
&[\delta v_2^\delta(t) - w_2^\delta(t)] \cdot \frac{\partial H^t}{\partial x_2}(\lambda z_1^\delta(t) + (1-\lambda) z_2^\delta(t)) = 0.
\end{aligned} \tag{9}$$

By (ii) in Lemma 3.1, it holds for every $i = 1, 2$ that

$$w_i^\delta(t) - \delta v_i^\delta(t) = \frac{(1-\delta) v_i^\delta(t)}{p_i}. \tag{10}$$

It follows from (9) and (10) that

$$\frac{v_1^\delta(t)}{p_1} \cdot \frac{\partial H^t}{\partial x_1}(\lambda z_1^\delta(t) + (1-\lambda) z_2^\delta(t)) = \frac{v_2^\delta(t)}{p_2} \cdot \frac{\partial H^t}{\partial x_2}(\lambda z_1^\delta(t) + (1-\lambda) z_2^\delta(t)). \tag{11}$$

Since $\{v^\delta(t) = (v_1^\delta(t), v_2^\delta(t))\}$ is a sequence in the compact set $U(t) \cap R_+$, it has some converging subsequence as δ goes to one. Let $v^*(t)$ be any limit of a subsequence of $\{v^\delta(t)\}$. It follows from (10) that

$$\lim_{\delta \to 1} v^\delta(t) = \lim_{\delta \to 1} z_1^\delta(t) = \lim_{\delta \to 1} z_2^\delta(t) = v^*(t).$$

Thus, by taking $\delta \to 1$ in (8) and (11), we obtain

$$H(v^*(t)) = 0 \quad \text{and} \quad \frac{v_1^*(t)}{p_1} \cdot \frac{\partial H}{\partial x_1}(v^*(t)) = \frac{v_2^*(t)}{p_2} \cdot \frac{\partial H}{\partial x_2}(v^*(t)).$$

Under Assumption 2.1, these conditions show that the limit $v^*(t)$ is the Nash bargaining solution with weights $p = (p_1, p_2)$ for the feasible set $U(t)$. Thus, the sequence $\{x_i^\delta\}$ of mechanisms proposed by every player $i = 1, 2$ converges to the ex post Nash bargaining solution with $p = (p_1, p_2)$ as δ goes to one. Q.E.D.

Proof of Lemma 5.1. By Proposition 3.1, proposals of all types of every player are accepted in the initial round in (σ, μ). If the equilibrium (σ, μ) satisfies property (ii) in the theorem, then the proof is complete. Suppose that this is not the case. Then for some player i, say $i = 1$, different types propose different mechanisms. Specifically, assume that there exist some partition $(T_1^1, \cdots, T_1^m)$ of T_1 and different mechanisms $x_1^1, \cdots, x_1^m$ such that all types of T_1^j propose x_1^j for each $j (= 1, \cdots, m)$.

We construct a mechanism $x_1^* \in M$ such that for every $t = (t_1, t_2) \in T$,

$$x_1^*(t) = x_1^j(t) \ \text{ if } \ t_1 \in T_1^j.$$

By construction, x_1^* is equal to the mechanism generated by (σ, μ). For player 2, we construct a mechanism x_2^* in the same way as x_1^*.

We define (σ', μ') according to the following rules:

(E1) If player 1 becomes a proposer, then he proposes x_1^* independent of his type.

(E2) Player 2 accepts proposal x_1^*, independent of his type. Each type t_2 for player 2 has the posterior belief $\pi(t_1|t_2)$ about the type of player 1, receiving proposal x_1^*.

(E3) If player 2 rejects x_1^* (off-play of σ'), then play is restarted according to (E1) or (E4) in the next round, depending on who becomes a proposer.

(E4) If player 2 becomes a proposer, then he proposes x_2^* independent of his type. The response and belief of player 1 receiving x_2^* are defined in the same way as for (E2). The same rule as in (E3) is applied.

(E5) Except for the rule above, let $\sigma = \sigma'$ and $\mu = \mu'$.

Clearly, (σ, μ) and (σ', μ') are outcome-equivalent, and (σ', μ') is stationary and satisfies IIT. Let $v(t) = (v_1(t), v_2(t))$ denote the conditional expected payoffs for players given a type profile t for both σ and σ'

We show that every type t_2 for player 2 optimally accepts x_1^* in (σ', μ'). If responder 2 accepts proposal x_1^* at t_2, then he receives the conditional expected payoff

$$\sum_{t \in T(t_2)} \pi(t_1|t_2) u_2(x_1^*(t), t) = \sum_{k=1}^{m} \sum_{t_1 \in T_1^k} \pi(t_1|t_2) u_2(x_1^k(t), t), \qquad (12)$$

where $t = (t_1, t_2)$.

Since the type partition for player 1, $(T_1^1, \cdots, T_1^m)$, is revealed on the play of (σ, μ), the sequential rationality of (σ, μ) means that for every $k = 1, \cdots, m$,

$$\sum_{t_1 \in T_1^k} \pi(t_1|T_1^k, t_2) u_2(x_1^k(t), t) \;\; \geq \;\; \sum_{t_1 \in T_1^k} \pi(t_1|T_1^k, t_2) \delta v_2(t), \qquad (13)$$

where $\pi(t_1|T_1^k, t_2) = \pi(t_1, t_2) / \sum_{t_1' \in T_1^k} \pi(t_1', t_2)$. From (12) and (13) it is evident that

$$\sum_{t_1 \in T_1} \pi(t_1|t_2) u_2(x_1^*(t), t) \geq \sum_{t_1 \in T_1} \pi(t_1|t_2) \delta v_2(t).$$

Therefore, it is optimal for every type t_2 for player 2 to accept x_1^*.

Since every type of player 1 in each T_1^k is indifferent to whether he proposes x_1^* or x_1^k (in σ), the sequential rationality of σ guarantees that x_1^* is the optimal proposal for him in σ'. The same argument as above holds true when player 2 is selected as a proposer. The sequential rationality of (σ', μ') at every other information set is trivially satisfied since $\sigma = \sigma'$ and $\mu = \mu'$ there according to (E5). Thus, (σ', μ') is a sequential equilibrium of Γ. Q.E.D.

References

Ausubel, L.M., P. Cramton and R.J. Deneckere (2002): "Bargaining with Incomplete Information," in *Handbook of Game Theory with Economic Applications,* Volume 3, ed. by R. Aumann and S. Hart, North-Holland, New York.

Baron, D. and J. Ferejohn (1989): "Bargaining in Legislatures," *American Political Science Review* 83, 1181–1206.

Binmore, K. A. Rubinstein and A. Wolinsky (1986): "The Nash Bargaining Solution in Economic Modelling,," *RAND Journal of Economics* 17, 176–188.

Chatterjee, K. and L. Samuelson (1987): "Bargaining with Two-sided Incomplete Information: an Infinite Horizon Model with Alternating Offers," *Review of Economic Studies* 65, 175–192.

Cho, I. and D. Kreps (1987): "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics* 102, 179–221.

de Clippel, G. and E. Minelli (2004): "Two-Person Bargaining with Verifiable Information," *Journal of Mathematical Economics* 40, 799–813.

Farrell, J. (1993): "Meaning and Credibility in Cheap-Talk Games," *Games and Economic Behavior* 5, 514–531.

Fudenberg, D. and J. Tirole (1983): "Sequential Bargaining with Incomplete Information," *Review of Economic Studies* 50, 221–247.

Grossman, S. J. and M. Perry (1986a): "Perfect Sequential Equilibrium," *Journal of Economic Theory* 39, 97–119.

Grossman, S. J. and M. Perry (1986b): "Sequential Bargaining under Asymmetric Information," *Journal of Economic Theory* 39, 120–154.

Harsanyi, J. C. (1968): "Games with Incomplete Information Played by "Bayesian" Players, Part II" *Management Science* 14, 320–334.

Harsanyi, J. C. and R. Selten (1972): "A Generalized Nash Solution for Two-Person Bargaining Games with Incomplete Information," *Management Science* 18, 80–106.

Kreps, D. and R. Wilson (1982): "Sequential Equilibria," *Econometrica* 50, 863–894.

Maskin, E. and J. Tirole (1990): "The Principal–Agent Relationship with an Informed Principal: the Case of Private Values," *Econometrica* 58, 379–409.

Maskin, E. and J. Tirole (1992): "The Principal–Agent Relationship with an Informed Principal, II: Common Values," *Econometrica* 60, 1–42.

Myerson, R. B. (1979): "Incentive Compatibility and the Bargaining Problem," *Econometrica* 47, 61–73.

Myerson, R. B. (1983): "Mechanism Design by an Informed Principal," *Econometrica* 51, 1767–1797.

Myerson, R. B. (1984): "Two-person Bargaining Problems with Incomplete Information," *Econometrica* 52, 461–487.

Mylovanov, T. and T. Tröger (2012): "Informed-Principal Problems in Environments with Generalized Private Values," *Theoretical Economics* 7, 465–488.

Nash, J. F. (1950): "The Bargaining Problem," *Econometrica* 18, 155–162.

Okada, A. (1996): "A Noncooperative Coalitional Bargaining Game with Random Proposers," *Games and Economic Behavior* 16, 97–108.

Okada, A. (2010): "The Nash Bargaining Solution in General n-Person Cooperative Games," *Journal of Economic Theory* 145, 2356–2379.

Okada, A. (2012): "Non-cooperative Bargaining and the Incomplete Informational Core," *Journal of Economic Theory* 147, 1165–1190.

Osborne, M. J. and A. Rubinstein (1990): *Bargaining and Markets,* Academic Press, New York.

Rubinstein, A. (1982): "Perfect Equilibrium in a Bargaining Model," *Econometrica* 50, 97–109.

Rubinstein, A. (1985): "A Bargaining Model with Incomplete Information about Time Preference," *Econometrica* 53, 1151–1172.

Sutton, J. (1986) "Non-cooperative Bargaining Theory: An Introduction," *Review of Economic Studies* 53, 709-724.

U.S. Department of State, New Strategic Arms Reduction Treaty (New START) at http://www.state.gov/documents/organization/140035.pdf, accessed on August 9, 2014.

Wilson R. (1978): "Information, Efficiency and the Core of an Economy," *Econometrica* 46, 807–816.