# The Nash Bargaining Solution in General $n$-Person Cooperative Games*

Akira Okada†

Hitotsubashi University

**Abstract**

We present a noncooperative foundation for the Nash bargaining solution for an $n$-person cooperative game in strategic form. The Nash bargaining solution should be immune to any coalitional deviations. Our noncooperative approach yields a new core concept, called the *Nash core*, for a cooperative game based on a consistency principle. We prove that the Nash bargaining solution can be supported (in every subgame) by a stationary subgame perfect equilibrium of the bargaining game if and only if the Nash bargaining solution belongs to the Nash core.

*JEL classification:* C71, C72, C78

Key words: Nash bargaining solution, noncooperative bargaining, coalition, externality, core, $n$-person cooperative games

*I am grateful to an associate editor and three anonymous referees for helpful comments and suggestions. I would also like to thank Tatsuro Ichiishi, Eric Maskin, Maria Montero and Debraj Ray for thoughtful discussions. Financial support from the Japan Society for the Promotion of Science under grant No.(A)16203011, No.(S)20223001 and the Matsushita International Foundation 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

The purpose of this paper is to develop a noncooperative game approach to a general $n$-person cooperative game. Specifically, we present a noncooperative foundation for the Nash bargaining solution in the game. A general cooperative game describes an economic situation in which $n$ individuals can communicate and form coalitions, which are enforceable and within which cooperation may have external effects on the utility of individuals outside the coalition. Individuals' utilities might not be transferable. The game covers a wide range of multilateral cooperation problems, including the following: production economy with externality, cartels by oligopolistic firms, public goods provision, environmental pollution, and international alliances. In this paper, a general cooperative game is described by an $n$-person game in strategic form.

In game theory, there have been two different approaches to the multilateral cooperation problem. One is the *cooperative game approach* initiated by the classic work of von Neumann and Morgenstern [33]. They reduce the $n$-person game in strategic form to its coalitional form (also called the characteristic function form), using as a basis the zero-sum two-person game played by coalitions. The von Neumann-Morgenstern approach may be regarded as a two-stage procedure. First, by using the minimax solution of a zero-sum two-person game between one coalition and its complementary coalition, one defines the *characteristic function* of a game that assigns to each coalition the set of payoff vectors that the coalition can "ensure" its members. Once the characteristic function has been constructed, cooperative solutions can be applied to investigate players' bargaining behavior. The *core* of a cooperative game is defined as the set of payoff allocations upon which no coalitions can improve by themselves. Two well-known core concepts in cooperative games with externality are the $\alpha$-core and the $\beta$-core (Aumann [1]). The $\alpha$-core corresponds to the characteristic function constructed by the maxmin value and the $\beta$-core corresponds to that constructed by the minimax value.

The von Neumann and Morgenstern theory (and thus the $\alpha$-core and $\beta$-core theory) assumes that when players form a coalition, they expect that the complementary coalition will react by damaging them in the worst way possible. This assumption about coalitional behavior has been often criticized in the literature on the ground that it allows threats by the complementary coalition that are not credible (Scarf [29], for example).[1]

The other approach is the *noncooperative game approach* initiated by Nash [21, 22] and called the Nash program. Nash proposed to study cooperative games on the basis of their reduction to noncooperative games by modelling pre-play negotiations as moves in a noncooperative bargaining game. Analyzing an equilibrium point[2] of the noncooperative bargaining game, one can explain coalitional behavior as the result of individual players' payoff maximization. Nash [22] presented a noncooperative foundation for his bargaining solution of a two-person cooperative game, which was presented by a set of axioms in his initial work (Nash [21]). An obvious restriction of Nash's work is that it covers only two-person games. To date, the noncooperative game approach to general $n$-person cooperative games has not been developed fully in the literature.

Given that individuals can form coalitions freely in a cooperative game, the Nash bargaining solution should be immune to any coalitional deviation. This claim may suggest that the Nash solution should satisfy a core stability. A critical question then arises: how should one define the core without allowing threats by players that are less than credible? Suppose that all $n$ players agree to the Nash bargaining solution in a game. If any coalition of players deviates from the Nash solution, then all other players have their own

[1] von Neumann and Morgenstern themselves point out a difficulty in the characteristic function approach. They write, "Now it would seem that the weakness of our present theory lies in the necessity to proceed in two stages: To produce a solution of the zero-sum two-person game first and then, by using this solution, to define a characteristic function in order to be able to produce a solution of the general $n$-person game, based on the characteristic function [33, p. 608].

[2] In what follows, a Nash equilibrium of a game will be simply called an equilibrium whenever no confusion arises.

bargaining problem of how to react to it. If one holds that the Nash bargaining theory should be applied to *every* negotiation problem, it should be the case that the remaining players react to the coalitional deviation according to the Nash bargaining solution. In other words, the Nash bargaining solution of a cooperative game must belong to a variant of the core of the game, in the sense that no coalition can improve upon it, anticipating the Nash bargaining solution behavior by the complementary coalition. We will call this new type of core for a cooperative game in strategic form the *Nash core*.

The notion of the Nash core is supported by the following argument for the consistency for a cooperative solution (the Nash bargaining solution in our case). Suppose that a cooperative solution is accepted as the standard of behavior in a game. Given that any coalition of players can be formed freely, the cooperative solution should be stable against any coalitional deviation. When some coalition deviates from the solution, the behavior of other players outside the coalition should be governed by the same standard of behavior (with no incentive to deviate). This argument for the consistency of standards of behavior leads naturally to the requirement that the Nash bargaining solution should belong to the Nash core. Given that the bargaining problem faced by players outside a coalition may be modelled as a subgame in a whole process of negotiations, it will be noted that the Nash core is closely linked to the notion of a subgame perfect equilibrium, which imposes an equilibrium on every subgame. Indeed, we will show that the reaction of the complementary coalition according to the Nash bargaining solution is not an ad hoc assumption but is a part of a subgame perfect equilibrium of the bargaining game in our noncooperative approach.

In this paper, we present a noncooperative bargaining model for a general $n$-person cooperative game that is based on a random-proposer model (Okada [23]) that is a generalization of the Rubinstein's [28] alternating offers model. In the model, a proposer is selected according to some predetermined probability distribution among active players. A proposal is a pair composed of a coalition and a jointly mixed action for its members. The proposal is accepted by unanimous consent among the members. Once a coalition forms, all members are bound to choose the agreed actions. If a proposal is rejected, there is a small probability that negotiations end, in which case all individuals play the game noncooperatively. Otherwise, the same process is repeated. We characterize the stationary subgame perfect equilibrium (SSPE) of the bargaining game when the probability of negotiation failure is sufficiently small. In particular, we are mainly concerned with the conditions under which all players agree to the Nash bargaining solution of the game. For this reason, our analysis focuses on an SSPE with the efficiency property that all active players cooperate in every round of the bargaining game. Such an equilibrium is called *totally efficient*.

The main results of the paper are summarized as follows. We first prove that if players form the largest coalition in an SSPE, their agreement should be the generalized (asymmetric) Nash bargaining solution, regardless of who proposed the agreement. The weights of players are determined by the probability distribution by which the proposers are selected, and the disagreement point is given by an equilibrium of the game. This result implies that, when one subcoalition is formed (off the equilibrium path) in a totally efficient SSPE, the complementary coalition reacts with the Nash bargaining solution for its own negotiation problem. Given an equilibrium as the disagreement point, a totally efficient SSPE exists uniquely, if at all, and all players agree to the Nash bargaining solution. We then prove that a totally efficient SSPE exists if and only if the Nash bargaining solution is in the Nash core of the game.[3] In this sense, we provide a bridge that connects the Nash equilibrium, the Nash bargaining solution, and the core in classical game theory.

Recently, several authors have refined the core solution in a cooperative game with externality so that it can eliminate incredible threats by players

[3]The if-part is proven using a technical condition that the interior of the (strict) Nash core is nonempty.

outside coalitions. Huang and Sjöström [14] and Kóczy [17] define the notion of a recursive core (abbreviated by 'r-core') by a theory of a solution's consistency that is similar to that of the Nash core. Roughly, the recursive core of a cooperative game with externality[4] is a variant of the core defined by the assumption that when a coalition forms, its members predict an outcome in a core of the "reduced game" composed of other players. Given that a solution's consistency requires that the same arguments be applied to the "core" of the reduced game, its definition is recursive. Chander and Tulkens [6] present another refinement of the core (called the $\gamma$-core) under the assumption that when a coalition forms, players outside the coalition adopt only individually best replies, which results in a Nash equilibrium among the coalition and the remaining players. Their concept does not satisfy consistency.

Although the Nash core and the recursive core are based on a general idea of consistency, it may be useful to make some remarks about differences in the two approaches. First, the works of Huang and Sjöström [14] and Kóczy [17] are motivated by a cooperative game approach with the aim of refining the core solution in a consistent manner. In contrast, our approach is a noncooperative one. The Nash core is a criterion for the Nash bargaining solution's being sustained as an SSPE of a noncooperative bargaining game. Huang and Sjöström [15] and Kóczy [16] provide a noncooperative implementation of the recursive core. Their bargaining models will be discussed in Section 6. Second, the result of the random proposer model is not limited to the Nash bargaining solution. It is well-known that a cooperative game does not always have a core. This may be true in the cases of the recursive core and the Nash core, as well. When the Nash core does not exist, the random proposer model yields an equilibrium outcome other than the Nash bargaining solution. An example of such a case is a majority voting game, which has been well-studied in the literature on legislative bargaining. In their seminal paper, Baron and Fere-

[4]Huang and Sjöström [14] employ a strategic form game and Kóczy [17] uses a partition function form game. The definition of the recursive core is essentially the same in either form.

john [2] characterize a unique SSPE of the majority voting game in which all minimal winning coalitions have the same probability of forming. Third, the core concept is a set-valued solution, unlike the Nash bargaining solution. To define the recursive core, one needs to make a behavioral assumption about how players outside a coalition react within the (recursive) core of the reduced game. Subject to the consistency constraint, Kóczy [17] presents the pessimistic recursive core and the optimistic recursive core in the same manner as the $\alpha$-core and the $\beta$-core, respectively.[5] In the pessimistic recursive core, members in a coalition predict that other players will behave towards them in the worst possible way under the constraint that their behavior is in the core of the reduced game. In a noncooperative approach, such a behavioral assumption is not needed. Any belief of the members, either pessimistic or optimistic, should be a part of a noncooperative equilibrium of a bargaining game.

The remainder of the paper is organized as follows. Section 2 provides definitions and notations. Section 3 presents a noncooperative bargaining model for an $n$-person cooperative game in strategic form. Section 4 states the main theorems. The proofs are given in Section 5. Section 6 discusses the results. Section 7 concludes.

# 2 Definitions and Notations

For a finite set $N$ with $n$ elements, let $R^N$ denote the $n$-dimensional Euclidean space with coordinates indexed by the elements of $N$. Any point in $R^N$ is denoted by $x = (x_i)_{i \in N}$, and also by $x = (x_1, x_2, \cdots, x_n)$ when $N$ is indexed as $\{1, 2, \cdots, n\}$. For $i \in N$ and $x = (x_i)_{i \in N} \in R^n$, $x_{-i}$ denotes the $(n-1)$-dimensional vector constructed from $x$ by deleting the $i$-th coordinate $x_i$. The point $x$ is sometimes written as $(x_i, x_{-i})$. For $S \subset N$, $R^S$ denotes the subspace

[5] Huang and Sjöström [14] consider the pessimistic recursive core. Shenoy [32] employs the optimistic approach in his core stability.

of $R^N$ spanned by the axes corresponding to elements in $S$. For a finite set $T$, the notation $\Delta(T)$ denotes the set of all probability distributions on $T$.

An *n-person cooperative game in strategic form* is defined by a triplet $G = (N, \{A_i\}_{i\in N}, \{u_i\}_{i\in N})$ where $N = \{1, 2, \cdots, n\}$ is the set of players and each $A_i$ $(i \in N)$ is a finite set of player $i$'s actions.[6] The Cartesian product $A = \Pi_{i\in N} A_i$ is the set of all action profiles $a = (a_1, \cdots, a_n)$. Player $i$'s payoff function $u_i$ is a real-valued function on $A$. A probability distribution on $A_i$ is called a mixed action for player $i$. A subset $S$ of $N$ is called a *coalition*. For a coalition $S$, let $A_S = \Pi_{i\in S} A_i$ be the set of action profiles $a_S = (a_i)_{i\in S}$ for all members of $S$. A *correlated action* $p_S$ of coalition $S$ is a probability distribution on $A_S$. The set of all correlated actions for the coalition $S$ is given by $\Delta(A_S)$. The idea of a correlated action is that all members in a coalition choose their actions jointly according to the corresponding probability distribution. In the cooperative game $G$, it is assumed that any coalition $S$ can make an enforceable agreement to employ any correlated action if all members agree to it.[7]

A *coalition structure* $\pi = [S_1, \cdots, S_m]$ on $N$ is defined by a partition of $N$, that is, a class of subsets of $N$ such that $N = S_1 \cup \cdots \cup S_m$ and every two $S_i$ and $S_j$ are disjoint. For a coalition structure $\pi = [S_1, \cdots, S_m]$ on $N$, an element $p^\pi = [p_{S_1}, \cdots, p_{S_m}]$ in $\Pi_{j=1}^m \Delta(A_{S_j})$ is called a *correlated action profile* for the coalition structure $\pi$. When a correlated action profile $p^\pi$ for $\pi = [S_1, \cdots, S_m]$ is employed, each player $i \in N$ obtains the expected payoff

$$u_i(p^\pi) = \sum_{a_{S_1}\in A_{S_1}} \cdots \sum_{a_{S_m}\in A_{S_m}} \Pi_{j=1}^m p_{S_j}(a_{S_j}) \cdot u_i(a_{S_1}, \cdots, a_{S_m}) \tag{1}$$

where $p_{S_j}(a_{S_j})$ $(j = 1, \cdots, m)$ is the probability that the correlated action $p_{S_j}$ of coalition $S_j$ assigns to an action profile $a_{S_j} \in A_{S_j}$. Given a coalition

[6]In this paper, we distinguish between "action" and "strategy", because we consider a sequential bargaining game in extensive form based on the game $G$ in strategic form.

[7]It suffices to assume the enforceability of pure actions only when the set $A_i$ of every player $i$'s pure actions is an infinite set and the set of all feasible payoff vectors, $F_S = \{(u_i(a_S, a_{N-S}))_{i\in S} | a_S \in A_S\}$, for coalition $S$ given an action profile $a_{N-S}$ for the complementary coalition $N - S$ is a compact and convex set.

structure $\pi = [S_1, \ldots, S_m]$ on $N$, we define

$$F(G, \pi) = \{(u_1(p^\pi), \cdots, u_n(p^\pi)) \in R^N \mid p^\pi \in \Pi_{j=1}^m \Delta(A_{S_j})\}. \tag{2}$$

The set $F(G, \pi)$ consists of all expected payoff vectors for $n$ players attained by correlated action profiles for $\pi$. When $\pi$ consists only of the grand coalition $N$, that is, $\pi = [N]$, $F(G, [N])$ is simply denoted by $F(G)$. We call $F(G)$ the *feasible set* of the cooperative game $G$. The feasible set $F(G)$ represents the set of all expected payoff vectors of $n$ players when they form the grand coalition $N$. The set $F(G)$ is a polyhedral compact convex subset of $R^N$ and $F(G) \supset F(G, \pi)$ for every coalition structure $\pi$ on $N$. The set $F(G, \pi)$ is not necessarily convex.

We define the Pareto frontier of the feasible set $F(G)$, following Harsanyi [11]. The upper-right boundary $H$ of $F(G)$ is defined as the set of points in $F(G)$ undominated (in the weak sense of Pareto) by any point in $F(G)$. With abuse of notation, we denote the equation of $H$ as@

$$H(x_1, \cdots, x_n) = 0$$

where $H$ is a function on the feasible set $F(G)$. With no loss of generality, we assume that $H(x) \geq 0$ for all $x \in F(G)$. In addition, for simplicity of the analysis, we assume:

**Assumption 1.**

(i) $H$ is a concave and differentiable function and the first derivatives of $H$ with respect to $x_1, \cdots, x_n$ satisfy

$$\frac{\partial H}{\partial x_1} \leq 0, \cdots, \frac{\partial H}{\partial x_n} \leq 0$$

(the equality may hold at most at the end points of the upper-right boundary $H$).

(ii) $F(G)$ has the full dimension $n$.

(iii) The strategic-form game $G$ has an equilibrium (in mixed actions) whose payoff vector $d = (d_1, \cdots, d_n) \in F(G)$ has the property that the boundary of the set $F_d(G) \equiv \{x \in F(G) \mid x_i \geq d_i \text{ for all } i \in N\}$, other than $n$ hyperplanes $x_i = d_i$ $(i \in N)$, is a subset of $H$.

Differentiability assumption (i) causes no loss of generality. We can easily extend our results to the non-differentiable case, given that the piecewise linear function of the upper-right boundary $H$ can constitute the limit of differentiable functions. The fact that all the first derivatives $\frac{\partial H}{\partial x_i}$ have the same sign implies that the variables $x_i$ and $x_j$ are mutually strictly decreasing functions of each other on the upper-right boundary $H$. For each $i \in N$, let $F_{-i}(G)$ denote the projection of $F(G)$ over $R^{N-\{i\}}$. For every $x_{-i} \in F_{-i}(G)$, we define $h_i(x_{-i}) = \max\{x_i \mid (x_i, x_{-i}) \in F(G)\}$. By assumption (i) and the convexity of $F(G)$, $h_i$ is a differential concave function over $F_{-i}(G)$. $h_i(x_{-i})$ is the maximum payoff that player $i$ can receive in the feasible set $F(G)$ while all other players' payoffs are fixed at $x_{-i}$. Assumptions (ii) and (iii) are technical. Assumption (iii) guarantees that for all $x \in F_d(G)$ the point $(h_i(x_{-i}), x_{-i})$ is located on the upper-right boundary $H$ of $F(G)$.

In the rest of this section, we introduce several notions of cooperative game theory. Since the classic work of von Neumann and Morgenstern [33], the *characteristic function* approach has been employed in cooperative game theory to study the problem of coalition formation and payoff distributions. The characteristic function of a cooperative game assigns to each coalition the set of payoff vectors that the coalition can "ensure" its members. For a strategic-form game, primarily the following two kinds of characteristic function have been studied (Aumann [1]). A coalition $S$ is said to be *$\alpha$-effective* for a payoff vector $x \in R^N$ if there exists $p_S \in \Delta(A_S)$ such that for any $p_{N-S} \in \Delta(A_{N-S})$, we have $u_i(p_S, p_{N-S}) \geq x_i$ for all $i \in S$. Let $v^\alpha(S)$ be the set of all payoff vectors for which $S$ is $\alpha$-effective. A coalition $S$ is said to be *$\beta$-effective* for $x \in R^N$ if for any $p_{N-S} \in \Delta(A_{N-S})$ there exists $p_S \in \Delta(A_S)$ such that

$u_i(p_S, p_{N-S}) \geq x_i$ for all $i \in S$. Similarly to $v^\alpha(S)$, let $v^\beta(S)$ be the set of all payoff vectors for which $S$ is $\beta$-effective. It is easily shown that $v^\alpha(S) \subset v^\beta(S)$ for every $S \subset N$. The functions $v^\alpha$ and $v^\beta$ are called the $\alpha$-characteristic function and the $\beta$-characteristic function, respectively.

Intuitively, $v^\alpha(S)$ is the set of all payoff vectors $x \in R^S$ such that coalition $S$ can guarantee all of its members at least the payoff $x$, independently of what the players in the complementary coalition $N-S$ do. On the other hand, $v^\beta(S)$ is the set of all payoff vectors $x \in R^S$ such that $N - S$ cannot prevent $S$ from getting at least $x$. In general, these two sets are different except for two-person games and for $n$-person games with transferable utility.

We now introduce two standard cooperative solution concepts: the core and the Nash bargaining solution.

**Definition 1.**

(1) Let $v = v^\alpha$ or $v^\beta$. A payoff vector $x \in R^N$ is said to *dominate* a payoff vector $y \in R^N$ *with respect to* $v$ if there exists some coalition $S$ of $N$ such that $x \in v(S)$ and $x_i > y_i$ for all $i \in S$

(2) The $\alpha$-*core* of a cooperative game $G$ is the set of payoff vectors $x \in F(G)$ that are not dominated by any other payoff vector in $F(G)$ with respect to $v^\alpha$. The $\beta$-*core* of $G$ is the set of payoff vectors $x \in F(G)$ that are not dominated by any other payoff vector in $F(G)$ with respect to $v^\beta$.

**Definition 2.** Let $\theta^N = (\theta_i^N)_{i \in N} \in \Delta(N)$, and $d^N = (d_i^N)_{i \in N} \in F(G)$. A correlated action $b^* \in \Delta(A_N)$ of $N$ is called the *(generalized) Nash bargaining solution* of $G$ if $b^*$ solves the maximization problem

$$\begin{array}{ll} \max & \displaystyle\sum_{i=1}^{n} \theta_i^N \cdot \log[u_i(p) - d_i^N] \\ \text{subject to} & (1)\ p \in \Delta(A_N) \\ & (2)\ u_i(p) \geq d_i^N \quad \text{for all} \quad i = 1, \cdots, n. \end{array}$$

Here, $\theta^N$ is called a *weight vector*, and $d^N$ the *disagreement point*. The Nash bargaining solution $b^*$ of $G$ with weight vector $\theta^N$ and disagreement point $d^N$ is denoted by $b^*(G, \theta^N, d^N)$ whenever the dependency on $G$, $\theta^N$ and $d^N$ should be emphasized. The payoff vector $u(b^*) = (u_i(b^*))_{i \in N}$ generated by the Nash bargaining solution $b^*$ is called the *Nash bargaining payoff*.

In negotiations, the grand coalition $N$ is not always formed. If the members of a coalition $S \subset N$ agree to choose a correlated action $p_S \in \Delta(A_S)$, all remaining players may continue their negotiations, given the agreement of the correlated action $p_S$ by $S$. The following game describes negotiations after some coalition has been formed.

**Definition 3.** Let $G$ be an $n$-person cooperative game in strategic form. For every coalition $S$ and every correlated action $p_S \in \Delta(A_S)$ of $S$, a *subgame* $G(p_S)$ of $G$ is defined to be the same game as $G$ except that all players in $S$ follow the correlated action $p_S$.[8]

The feasible set $F(G(p_S))$ of a subgame $G(p_S)$ can be defined in the same manner as the feasible set $F(G)$ of $G$. Note that the set of "active" players is $N - S$ in the subgame $G(p_S)$. The model of a subgame $G(p_S)$ of $G$ can describe a general situation in which several coalitions have formed, by considering $S$ as the union of these coalitions and $p^S$ as the correlated action of $S$ generated by the correlated actions that have been agreed in subcoalitions.

Our cooperative solution for a strategic-form game $G$ does not simply specify a feasible payoff (or a correlated action) for the grand coalition $N$. Rather, it is a *payoff configuration*, which specifies for *every* coalition $S$ of $N$ a feasible

[8]Here we should not confuse a subgame of $G$ with the standard notion of a subgame in an extensive-form game, although it turns out that every subgame of $G$ naturally corresponds to a subgame of a noncooperative bargaining game in extensive form introduced in Section 3.

payoff of $S$.[9] Given that the feasible payoff for the coalition $S$ depends on a correlated action of the complementary coalition $N-S$ in our set-up, a payoff configuration specifies a feasible payoff of $S$ for every correlated action $p^{N-S}$ of $N-S$. Formally, a payoff configuration for $G$ is defined as a function $\phi$ that assigns for every coalition $T$ and every correlated action $p_T$ of $T$ an element $\phi(p_T)$ in the feasible set $F(G(p_T))$ of the subgame $G(p_T)$. (Put $S = N - T$ in the discussion above). The vector $\phi(p_T)$ indicates the payoffs other players receive when coalition $T$ forms by agreeing to choose $p_T$. In the next section, we will see that a payoff configuration for $G$ can be derived naturally by a strategy profile for a noncooperative bargaining game of $G$.

We extend the Nash bargaining solution of $G$ to a solution configuration of $G$. Let $\theta$ be a function that assigns to each $S \subset N$ a weight vector $\theta^S \in \Delta(S)$ of its members. We call $\theta$ the *weight configuration* of $N$. Let $d$ be a function that assigns to every correlated action $p_S \in \Delta(A_S)$ of every coalition $S$ a payoff vector $d(p_S)$ in the feasible set $F(G(p_S))$ of the subgame $G(p_S)$. The vector $d(p_S)$ is interpreted as a disagreement point for negotiations within the complementary coalition $N-S$, given that $S$ employs the correlated action $p_S$. We call $d$ the *disagreement configuration* of $G$.

**Definition 4.** The *Nash bargaining solution configuration* $b^*$ of $G$ with weight configuration $\theta$ of $N$ and disagreement configuration $d$ is the function that assigns to every correlated action $p_S$ of every coalition $S$ the Nash bargaining solution $b^*(p_S) = b^*(G(p_S), \theta^{N-S}, d(p_S))$ of the subgame $G(p_S)$.[10] The payoff configuration of $G$ generated by $b^*$ is called the *Nash bargaining payoff configuration.*

[9]The formulation of a cooperative solution as a payoff configuration is employed in Hart's [13] axiomatization of the Harsanyi value for a cooperative game with non-transferable utility.

[10]For notational simplicity, we use the same symbol $b^*$ for the Nash bargaining solution configuration as for the Nash bargaining solution in Definition 2.

The characteristic function, which prescribes what a coalition can achieve by itself, has played a central role in cooperative game theory since von Neumann and Morgenstern [33]. A characteristic function of a strategic-form game assumes a certain behavior of a coalition $S$ and the complementary coalition $N - S$. In the $\alpha$-characteristic function, a coalition $S$, in attempting to improve its position, takes into account all strategic possibilities that are open to the complementary coalition $N - S$. The $\alpha$-characteristic function has been criticized (Scarf [29], for example) on the grounds that a coalition $S$ pays too much attention to threats by members of $N-S$ that may be harmful to themselves. Alternatively, one can argue that a counter-action of $N - S$ should be consistent with the members' utility maximization. From this point of view, by using the Nash bargaining solution configuration, we define a new notion of effectiveness for a cooperative game in strategic form that is weaker than $\alpha$-effectiveness.

**Definition 5.** Let $b^*$ be the Nash bargaining solution configuration of a cooperative game $G$ with weight configuration $\theta$ and disagreement configuration $d$.

(i) A coalition $S \subset N$ is said to be *Nash-effective* for a payoff vector $x \in R^n$ if there exists $p_S \in \Delta(A^S)$ such that

$$u_i(p_S, b^*(p_S)) \geq x_i \quad \text{for all} \quad i \in S \tag{3}$$

where $b^*(p_S) = b^*(G(p_S), \theta^{N-S}, d(p_S))$ is the Nash bargaining solution of the subgame $G(p_S)$ of $G$ assigned by $b^*$ under $\theta$ and $d$.

(ii) The *Nash characteristic function* $v^{Nash}$ of $G$ is the function that assigns to each coalition $S \subset N$ the set, denoted by $v^{Nash}(S)$, of all payoff vectors in $R^n$ for which $S$ is Nash-effective.

(iii) The *Nash core* of $G$ is the core of $G$ with respect to the Nash characteristic function $v^{Nash}$.

Nash effectiveness is based on the following idea. When a coalition $S$ chooses a correlated action $p_S$, it should consider a counter-action of the complementary coalition $N - S$ that is consistent with its members' payoff maximization. In other words, the coalition $S$ should consider only credible threats by $N - S$. A question remains: what is the outcome of the payoff-maximizing behavior of $N - S$? Given that the members of $N - S$ can negotiate their reaction to $S$, it is reasonable to assume that they agree to choose the Nash bargaining solution of their negotiation problem described by the subgame $G(p_S)$, because $S$ chooses a correlated action $p_S$. It is easily seen that Nash effectiveness is weaker than $\alpha$-effectiveness. Thus, the Nash core is a refinement of the $\alpha$-core, and it requires that a threat of the complementary coalition $N - S$ against $S$ be credible in the sense that the threat is consistent with the Nash bargaining theory.[11] As we remarked in the introduction, the Nash core may be empty, just like other solution concepts in the core family. The next example considers the possibility of emptiness.

Before we present a noncooperative bargaining model for the Nash bargaining solution in a general cooperative game, we illustrate the idea of the Nash core with the help of a three-person game.

**Example 1.** (a three-person prisoner's dilemma)

Consider the three-person game $G$ in strategic form that is given in Fig.1. The game can be interpreted as a prisoner's dilemma. Every player $i$ (=1, 2, 3) has two actions, $C_i$ (cooperate) and $D_i$ (defect). Every player $i$ has the dominant action $D_i$, and thus the game has a unique Nash equilibrium $(D_1, D_2, D_3)$. It can be seen that the action profile $(C_1, C_2, C_3)$ is the (symmetric) Nash bargaining solution of the cooperative game $G$ with disagreement point $(D_1, D_2, D_3)$.

[11] Chakrabarti [5] considers a refinement of the $\beta$-core that requires that a threat used by $N - S$ to all possible deviations by $S$ be a Pareto-undominated action of $N - S$.

|  | $C_2$ | $D_2$ |
|---|---|---|
| $C_1$ | 2,2,2 | 0,6,0 |
| $D_1$ | 6,0,0 | 3,3,-2 |

$C_3$

|  | $C_2$ | $D_2$ |
|---|---|---|
| $C_1$ | 0,0,6 | -2,3,3 |
| $D_1$ | 3,-2,3 | 1,1,1 |

$D_3$

Fig.1. A three-person prisoner's dilemma game

We will show that the Nash bargaining solution $(C_1, C_2, C_3)$ is in the Nash core of the cooperative game $G$. Before we construct the Nash characteristic function, we explain the basic idea of the Nash core. Suppose that a single player $i$, say $i = 3$, defects from the Nash bargaining solution. Then, players 1 and 2 negotiate about how to react to player 3's defection. Their strategic possibilities are described by the two-person game $G_{\{1,2\}}$ in Fig.2. In the game $G_{\{1,2\}}$, $(D_1, D_2)$ is a dominant equilibrium, and thus is the unique disagreement point. Given that $(D_1, D_2)$ is Pareto efficient in the game $G_{\{1,2\}}$, it is trivially the Nash bargaining solution of $G_{\{1,2\}}$. That is, players 1 and 2 agree to react to player 3's defection by playing $(D_1, D_2)$. Then, player 3's payoff decreases from 2 to 1. Player 3 is worse off by defecting. Next, suppose that any two players, say 1 and 2, defect jointly. Then, player 3 reacts to this coalitional defection by defecting herself, because $D_3$ is her optimal action to $(D_1, D_2)$. Then, the payoff of both players 1 and 2 decreases from 2 to 1. Players 1 and 2 are worse off by defecting jointly. Given that no coalition can improve upon the Nash bargaining solution $(C_1, C_2, C_3)$, this solution belongs to the Nash core.

|  | $C_2$ | $D_2$ |
|---|---|---|
| $C_1$ | 0, 0 | -2, 3 |
| $D_1$ | 3, -2 | 1, 1 |

Fig.2. A two-person game $G_{\{1,2\}}$ between players 1 and 2 when player 3 defects.

The Nash characteristic function of the cooperative game $G$ is constructed as follows. Suppose that player 3 employs a mixed action $p_3 = (p, 1-p)$ where $p$ $(0 \leq p \leq 1)$ is the probability of selecting $C_3$. Then, by the same argument as above, players 2 and 3 react to player 1 by employing the Nash bargaining solution $(D_2, D_3)$ of their own bargaining problem. Therefore, the Nash characteristic function $v^{Nash}(\{3\})$ for player 3 is given by

$$v^{Nash}(\{3\}) = \{w_3 \in R | \ \ w_3 \leq 1 - 3p \ \text{ for some } p, \ 0 \leq p \leq 1\}.$$

Given that $2 > 1 - 3p$ for any $p$ $(0 \leq p \leq 1)$, player 3 cannot improve upon the Nash bargaining solution $(C_1, C_2, C_3)$. The same result holds for $i = 1, 2$.

Next, suppose that players 1 and 2 jointly employ a correlated action $p^{12} = (p, q, r, 1 - p - q - r)$ where $p$ is the probability for an action profile $(C_1, C_2)$, $q$ the probability for $(C_1, D_2)$, and $r$ the probability for $(D_1, C_2)$. Given that player 3 chooses the dominant action $D_3$, the Nash characteristic function $v^{Nash}(\{1, 2\})$ for players 1 and 2 is given by

$$\begin{aligned} v^{Nash}(\{1,2\}) = \{(w_1, w_2) \in R^2 | \ \ & w_1 \leq 1 - p + 2q - 3r, w_2 \leq 1 - p - 3q + 2r \\ & \text{for some } p, q, r \text{ with } 0 \leq p, q, r \leq 1, \ 0 \leq p + q + r \leq 1\}. \end{aligned}$$

It is impossible for both $1 - p + 2q - 3r > 2$ and $1 - p - 3q + 2r > 2$ to hold simultaneously for some $p$, $q$ and $r$ with $0 \leq p + q + r \leq 1$. Therefore, coalition $\{1, 2\}$ cannot improve upon the Nash bargaining solution $(C_1, C_2, C_3)$. The same result holds for any other two-person coalition. Thus, $(C_1, C_2, C_3)$ belongs to the Nash core.

Finally, we remark that the Nash solution $(C_1, C_2, C_3)$ does not belong to the Nash core if the payoff vector for the action profile $(D_1, D_2, D_3)$ is changed from (1, 1, 1) to (-1, -1, -1) in Fig.1. In the new game, if player 3 defects, then players 1 and 2 agree to react with the Nash solution of their own negotiation problem with disagreement point (-1, -1), which means that they play $(C_1, D_2)$ and $(D_1, C_2)$ with equal probability. Then, player 3 receives payoff 3, higher than 2. Thus, player 3 can improve upon the Nash solution $(C_1, C_2, C_3)$ of $G$. In this new game, the Nash solution $(C_1, C_2, C_3)$ does not belong to the Nash core. In fact, since every player can guarantee payoff 3, the Nash core is empty.

# 3 A Noncooperative Bargaining Model

A bargaining model of an $n$-person cooperative game $G$ in strategic form is divided into two consecutive phases, (I) negotiation and (II) play. The negotiation phase consists of a (possibly) infinite sequence of bargaining rounds. In the second phase, all players who have agreed to form coalitions implement their agreed-upon correlated actions, and the other players choose their actions independently. Let $\theta$ be a weight configuration of $N$. The precise rules for bargaining are explained below.

(I) negotiation phase:

This phase has a (possibly) infinite number of bargaining rounds $t$ $(= 1, 2, \cdots)$. Let $N_t$ be the set of all "active" players who do not belong to any coalitions in round $t$. In the initial round, we put $N_1 = N$. At the beginning of each round $t$, a proposer is selected at random from the set $N_t$ of active players according to the probability distribution $\theta(N_t) \in \Delta(N_t)$ that the weight configuration $\theta$ assigns to $N_t$. The selected player $i \in N_t$ proposes a coalition $S$ with $i \in S \subset N_t$ and a correlated action $p_S \in \Delta(A_S)$ for $S$. All other members in $S$ either accept or reject the proposal sequentially according to a predetermined order over $N_t$. The order of responders does not affect the result in any critical way. If all responders accept the proposal $(S, p_S)$, then it is binding. Then, negotiation goes to the round $t + 1$ with $N_{t+1} = N_t - S$. The process is repeated with the probability distribution $\theta^{N_{t+1}} \in \Delta(N_{t+1})$.

If any one responder rejects the proposal, with probability $1-\varepsilon$ $(\varepsilon > 0)$, the negotiations continue in the round $t + 1$ with $N_{t+1} = N_t$. With probability $\varepsilon$, negotiations break down and the game goes to the next phase. The negotiation process ends when every player in $N$ joins some coalition.

(II) play phase:

In the second phase, all players actually play the game $G$. If players belong to coalitions, they are bound to implement their agreed-upon correlated actions. If there remain any players outside coalitions when negotiation stops, they may choose their individual (mixed) actions independently.[12]

The bargaining model above is denoted by $\Gamma^{\varepsilon,\theta}$. Formally, $\Gamma^{\varepsilon,\theta}$ is represented as an infinite-length extensive game with perfect information (that is, all players know all past actions when they make their choices). The rules of the game are commonly known to players. We also use a notation $\Gamma^{\theta}$ to describe the bargaining model where the probability $\varepsilon$ that negotiations will stop converges to 0.

A *(behavior) strategy* for player $i$ in $\Gamma^{\varepsilon,\theta}$ is defined according to the standard theory of extensive games. Let $h_i^t$ be a history of the game $\Gamma^{\varepsilon,\theta}$ when it is player $i$'s turn to move in round $t$ of the negotiation phase. The history $h_i^t$ consists of the sequence of all past actions in $\Gamma^{\varepsilon,\theta}$ before player $i$'s move in round $t$. Specifically, it describes proposers, proposals, and responses in all past rounds.[13] Similarly, let $\bar{h}$ be a whole history of the negotiation phase when the play phase starts. Roughly, a strategy $s_i$ of player $i$ in $\Gamma^{\varepsilon,\theta}$ is a function that assigns her action $s_i(h)$ to every possible history $h = h_i^t$ or $\bar{h}$. Specifically, player $i$'s action $s_i(h)$, $h = h_i^t$ or $\bar{h}$, is given as follows: (i) when player $i$ is a proposer in round $t$, $s_i(h_i^t)$ is a probability distribution (with finite support[14]) on the set of all possible proposals $(S, p_S)$ with $i \in S \subset N_t$ and $p_S \in \Delta(A_S)$, (ii) when a responder in round $t$, $s_i(h_i^t)$ is a probability distribution over {accept,

[12]When negotiation continues forever (the probability of which event is 0 as long as the stopping probability $\varepsilon$ of negotiations is positive), it is assumed for completeness of the model that all players who do not join any coalitions choose their actions independently.

[13]When player $i$ is a responder in round $t$, $h_i^t$ includes the proposer, the proposal, and all responses before player $i$ in the current round $t$, too.

[14]The assumption of finite support does not affect the result at all because any probability mixture, with finite or infinite support, of correlated actions $p_S \in \Delta(A_S)$ can be reduced to a single correlated action in $\Delta(A_S)$.

reject}, and (iii) in the play phase, $s_i(\bar{h})$ is player $i$'s mixed action in $\Delta(A_i)$ if she does not belong to any coalition; otherwise, she follows the agreed-upon correlated action.

Let $P$ denote the set of all correlated action profiles $p^\pi$ for all coalition structures $\pi$ of $N$. For a strategy profile $s = (s_1, \cdots, s_n)$ of players in $\Gamma^{\varepsilon,\theta}$, a probability distribution $\mu$ on $P$ (with finite support) is determined. Then, player $i$'s expected payoff for a strategy profile $s$ is given by

$$Eu_i(s) = \int_P u_i(p^\pi) d\mu, \tag{4}$$

where $u_i(p^\pi)$ is the expected payoff of player $i$ for the correlated action profile $p^\pi$. In what follows, the expected payoff $Eu_i(s)$ is denoted by $u_i(s)$, with abuse of notation, and the expected payoff is simply called 'payoff', whenever no confusion arises.

For every correlated action $p_S$ of every coalition $S$, let $\Gamma^{\varepsilon,\theta}(p_S)$ be the subgame of the extensive game $\Gamma^{\varepsilon,\theta}$ that starts after the agreement $(S, p_S)$ has been reached. In the same way as (4), a strategy profile $s = (s_1, \cdots, s_n)$ of players in $\Gamma^{\varepsilon,\theta}$ generates the expected payoff vector for players in the subgame $\Gamma^{\varepsilon,\theta}(p_S)$, which is an element of the feasible set $F(G(p_S))$ of the game $G(p_S)$. In this way, a strategy profile $s = (s_1, \cdots, s_n)$ in $\Gamma^{\varepsilon,\theta}$ naturally generates a payoff configuration for the cooperative game $G$.

The solution concept that we apply to the bargaining game $\Gamma^{\varepsilon,\theta}$ is a stationary subgame perfect equilibrium.

**Definition 6.** A strategy combination $s^* = (s_1^*, \cdots, s_n^*)$ of the game $\Gamma^{\varepsilon,\theta}$ is called a *stationary subgame perfect equilibrium (SSPE)* if $s^*$ is a subgame perfect equilibrium of $\Gamma^{\varepsilon,\theta}$ where every player $i$'s strategy $s_i^*$ is *stationary*, i.e., satisfies the property that the action $s_i^*(h)$ prescribed by $s_i^*$ to any history $h$ depends only on the collection of agreements, $(S_1, p_{S_1}), \cdots, (S_m, p_{S_m})$ which have been reached on $h$.[15] A payoff (configuration) generated by an SSPE is called an SSPE payoff (configuration).

Agreements of coalitions constitute a payoff-relevant history of negotiations. The SSPE requires that every player's equilibrium action should depend only on such payoff-relevant history. However, it should be emphasized that deviations from the equilibrium are allowed to be nonstationary. In the context of negotiations, this represents forgiveness - "let bygones be bygones." Players do not treat one another unfavourably even if they were treated so in past rounds of negotiations.

It is well-known that in a broad class of Rubinstein-type sequential multilateral bargaining games, including our game $\Gamma^{\varepsilon,\theta}$, there are many subgame perfect equilibria when the discount rate of future payoffs, or the probability that negotiations will break down, is very small (Osborne and Rubinstein [25]). Multiplicity of these equilibria holds even in the $n$-person pure bargaining game where no subcoalitions are allowed. It is mainly for this reason that the concept of an SSPE is employed in almost all studies of noncooperative multilateral bargaining models (Selten [31], Baron and Ferejohn [2], Perry and Reny [26] and Chatterjee et al. [7], among others). One possible justification for an SSPE is a focal-point argument. It is the simplest type of subgame perfect equilibrium, and thus it may be easier for players to coordinate their expectations on it (Baron and Kalai [3]). The SSPE is a natural reference point of the analysis in multilateral bargaining models.

In the literature on equilibrium selection in noncooperative games, SSPE is equivalent to subgame perfect equilibrium satisfying the condition of *subgame consistency* introduced by Harsanyi and Selten [12]. Subgame consistency in general extensive games requires that every player behave in the same way across "isomorphic" subgames. In the context of our bargaining game $\Gamma^{\varepsilon,\theta}$, all

[15] Precisely speaking, when player $i$ is a responder, her response surely depends on a current proposal and may depend on who a proposer is, and on how responders that preceded her have behaved in the same round.

subgames starting at the beginning of all rounds can be considered isomorphic as long as the same collections of agreements have been reached before, because if that is so, they will have identical game trees. In addition, an SSPE can be reformulated as a Markov-perfect equilibrium (Maskin and Tirole [18]) of $\Gamma^{\varepsilon,\theta}$ by taking the collection of agreements reached in past negotiations as a payoff-relevant state variable in each round.

The bargaining game $\Gamma^{\varepsilon,\theta}$ may suffer from two kinds of inefficiency. First, a proposal is rejected and negotiations break down with a positive probability. Breakdown of negotiations typically results in an inefficient outcome. Second is the failure of the grand coalition $N$ to form. It is known that the first kind of inefficiency may occur in the fixed-order model where an initial proposer is determined according to a fixed order over the player set and the first rejector becomes the next proposer (Chatterjee et al. [7]). Okada [23] proves that when utility is transferable, this is not the case in the random proposer model for a super-additive TU game. In the next section, it will be shown that this result can be extended to a cooperative game in strategic form. Specifically, we will prove that in every SSPE of $\Gamma^{\varepsilon,\theta}$, every player's proposal is accepted in the first round. This enables us to focus on the problem of inefficiency caused by the formation of subcoalitions.

**Definition 7.**

(i) An SSPE $s$ of $\Gamma^{\varepsilon,\theta}$ is called *efficient* if the grand coalition $N$ is formed in the initial round of the negotiation phase, independently of who the proposer is.

(ii) An SSPE $s$ of $\Gamma^{\varepsilon,\theta}$ is called *totally efficient* if the coalition of all active players (if any) are formed in every round of the negotiation phase, independently of history.

(iii) A *limit efficient* SSPE of $\Gamma^{\theta}$ is defined to be a limit of efficient SSPEs of $\Gamma^{\varepsilon,\theta}$ as $\varepsilon$ goes to zero. A *limit totally efficient* SSPE of $\Gamma^{\theta}$ is defined to be a limit of totally efficient SSPEs of $\Gamma^{\varepsilon,\theta}$ as $\varepsilon$ goes to zero.

In an efficient SSPE, the grand coalition $N$ is formed in the initial round of negotiations on the equilibrium path. A totally efficient SSPE satisfies the stronger property that the coalition of all active players is formed not only *on* the equilibrium path but also *off* it. In other words, the totally efficient SSPE of $\Gamma^{\varepsilon,\theta}$ induces an efficient SSPE on every subgame $\Gamma^{\varepsilon,\theta}(p_S)$ of $\Gamma^{\varepsilon,\theta}$ where $p_S$ is a correlated action of a coalition $S$, independently of whether it is reached by the equilibrium path or not. Obviously, a totally efficient SSPE of $\Gamma^{\varepsilon,\theta}$ is an efficient SSPE.

# 4 Theorems

The aim of our analysis is to characterize the limit-totally-efficient SSPE in the bargaining game $\Gamma^{\theta}$. In this section, we will state the main theorems. All proofs are given in the next section. The following proposition is useful to our analysis.

**Proposition 1**. (No delay) Let $s^*$ be an SSPE of $\Gamma^{\varepsilon,\theta}$. Then, for every $i \in N$, player $i$'s proposal is accepted in the initial round of the negotiation phase in $s^*$.

The proposition shows that there is no delay of agreement in the bargaining game $\Gamma^{\varepsilon,\theta}$. That is, some agreement of coalition is reached immediately on the equilibrium path. The bargaining rule of $\Gamma^{\varepsilon,\theta}$ that a proposer is selected at random in every round, is critical to this result. The theorem does not hold in the fixed-order model. Montero [20] shows the no-delay result of the random proposer model for a game in partition function form. In Proposition 1, we remark that the grand coalition is not necessarily formed.

We are now ready to state the main theorems.

**Theorem 1.** Let $v = (v_1, \cdots, v_n)$ be a limit-efficient SSPE payoff of $\Gamma^\theta$. Let $\theta^N$ be the weight vector for $N$ assigned by the weight configuration $\theta$. Then, $v$ is the Nash bargaining payoff of the cooperative game $G$ with weight vector $\theta^N$ and disagreement point $d = (d_1, \cdots, d_n)$. The disagreement point $d$ is given by an equilibrium payoff of $G$.

The theorem shows that when the probability $\varepsilon$ that negotiations will stop upon rejection is sufficiently small, players agree to the Nash bargaining solution in an efficient SSPE of $\Gamma^{\varepsilon,\theta}$. Two remarks are in order. First, the disagreement point of the Nash bargaining solution is given by a Nash equilibrium in the strategic-form game $G$. Unlike Nash's [22] optimal threat model, our bargaining game $\Gamma^{\varepsilon,\theta}$ (and $\Gamma^\theta$) does not allow players to commit themselves to incredible threats that will be implemented when negotiations fail. The SSPE of $\Gamma^{\varepsilon,\theta}$ prescribes that players should play an equilibrium of $G$ when negotiations break down. Secondly, the theorem shows that the weights of players for the Nash bargaining solution are determined by the likelihood that they will make proposals. The more likely it is that a player will be selected as a proposer, the greater her bargaining power. In the context of a two-person game, Binmore et al. [4] show other sources of asymmetry. They include different waiting times to make counter offers after rejection and different beliefs concerning the probability of breakdown.

With the help of Theorem 1, we characterize a limit-totally-efficient SSPE of $\Gamma^\theta$. By definition, a totally efficient SSPE of $\Gamma^\theta$ induces a totally efficient SSPE of every subgame $\Gamma^\theta(p_S)$ of it that starts after a coalition $S$ agrees to play a correlated action $p_S$. In other words, the members of the coalition $S$ should anticipate the totally efficient SSPE behavior of the complementary coalition. This observation leads naturally to the notions of Nash effectiveness and thus of the Nash core (Definition 5). A limit-totally-efficient SSPE payoff of $\Gamma^\theta$ must be in the Nash core. If not, there exists some coalition $S$ whose members can improve upon their SSPE payoffs by employing some correlated action (that will be counteracted by the Nash bargaining solution behavior of the complementary coalition $N - S$). Every member of $S$ has an incentive to propose such a coalitional deviation (when selected as a proposer) since all other members of $S$ accept it. This contradicts the SSPE property. The Nash bargaining solution configuration that defines the Nash core has a disagreement configuration $d$ that satisfies the following property:

**(A)** For every correlated action $p_S \in \Delta(A_S)$ of every coalition $S$, the disagreement configuration $d$ of $G$ assigns an equilibrium payoff of the subgame $G(p_S)$ of $G$.

**Theorem 2.** Let $\phi^*$ be the payoff configuration generated by a limit-totally-efficient SSPE $s^*$ of $\Gamma^\theta$. Then
(i) $\phi^*$ is the Nash bargaining payoff configuration with weight configuration $\theta$ and disagreement configuration $d$ that satisfies (A), and
(ii) for every $S \subset N$ and every $p_S \in \Delta(A_S)$, the payoff $\phi^*(p_S) \in F(G(p_S))$ assigned by $\phi^*$ belongs to the Nash core of the subgame $G(p_S)$ that is defined by the Nash bargaining solution configuration associated with $\theta$ and $d$.[16]

It follows from Theorem 2 that when the probability $\varepsilon$ that negotiations break down is very small, a totally efficient SSPE payoff is the Nash bargaining payoff with weight vector $\theta^N$, and moreover that it belongs to the Nash core of $G$. Given that the totally efficient SSPE of $\Gamma^{\varepsilon,\theta}$ has the *subgame property*, namely that it induces a totally efficient SSPE on *every* subgame $\Gamma^{\varepsilon,\theta}(p_S)$ of $\Gamma^{\varepsilon,\theta}$, the property above of the totally efficient SSPE payoff should be true for every subgame $\Gamma^{\varepsilon,\theta}(p_S)$.

To understand condition (ii) of Theorem 2, we discuss what it means in a special case of a transferable utility game $(N, v)$ in characteristic function

[16]When $S$ is the empty set $\emptyset$, the subgame $G(p^S)$ means the whole game $G$.

form where the characteristic function $v$ assigns a real value $v(S)$ to every coalition $S$ of $N$. For a coalition $S$, the restriction of $v$ to $S$ is denoted by $v_S$. A subgame $G(p_{N-S})$ of $G$ simply corresponds to a transferable utility game $(S, v_S)$ with player set $S$. The (symmetric) Nash bargaining solution of $(S, v_S)$ with disagreement point $v(\{i\}) = 0$ for all $i \in S$ is given by the equal payoff vector $(1/|S|, \cdots, 1/|S|)$ where $|S|$ denotes the number of members in $S$. Given that the value $v(S)$ of coalition $S$ is independent of the action taken by the complementary coalition $N - S$, the Nash core of the game $(S, v_S)$ is equal to the usual core. The Nash bargaining solution $(1/|S|, \cdots, 1/|S|)$ belongs to the core of $(S, v_S)$ if and only if $v(S)/|S| \geq v(T)/|T|$ for all subcoalitions $T$ of $S$. Therefore, for equal weights, the condition (ii) of Theorem 2 reduces to a simple condition: $v(S)/|S| \geq v(T)/|T|$ for all coalitions $S$ and $T$ of $N$ with $T \subset S$. We proved in Okada [23, Theorem 3] that the converse of Theorem 2 holds true for a transferable utility game in characteristic function form. The last theorem shows that this result can be extended to a general cooperative game $G$ in strategic form (under certain technical conditions).

**Theorem 3.** Let $\phi^*$ be the Nash bargaining payoff configuration of a cooperative game $G$ with weight configuration $\theta$ and disagreement configuration $d$ that satisfies (A). If $\phi^*$ satisfies

**(B)** for every $S \subset N$ and every $p_S \in \Delta(A_S)$, the payoff $\phi^*(p_S) \in F(G(p_S))$ assigned by $\phi^*$ belongs to the interior of the strict Nash core[17] of subgame $G(p_S)$ relative to the upper-right boundary of the feasible set $F(G(p_S))$,

then $\phi^*$ is a payoff configuration generated by a limit-totally-efficient SSPE of $\Gamma^\theta$.

Theorems 2 and 3 virtually show that the Nash bargaining solution of the

[17]The strict core is defined in the same manner as the core, except that the domination requires that no member of a coalition is ever worse off with at least one member being better-off. When utility is transferable, the core and the strict core coincide.

general cooperative game $G$ can be supported by the totally efficient SSPE of the bargaining model $\Gamma^{\theta}$, where the probability that negotiations fail is very small, if and only if the Nash bargaining payoff belongs to the Nash core. We remark that when the feasible set $F(G)$ (or $v^{Nash}(N)$) is large compared to $v^{Nash}(S)$ for all $S \neq N$, the Nash core becomes large, and thus condition (B) in Theorem 3 is more likely to hold.

# 5 Proofs

In this section, we will prove the results with the help of several lemmas.

**Lemma 1.** Let $s^* = (s_1^*, \cdots, s_n^*)$ be an SSPE of $\Gamma^{\varepsilon,\theta}$, and let $q^* = (q_1^*, \cdots, q_n^*)$ be a mixed action profile of $G$ that is employed by $s^*$ in the play phase of $\Gamma^{\varepsilon,\theta}$ when no agreements have been reached in the negotiation phase. Then $q^*$ is an equilibrium of $G$.

**Proof.** When no agreements have been reached in the negotiation phase, all players select their actions independently in the play phase, and thereafter the bargaining game $\Gamma^{\varepsilon,\theta}$ ends. This rule of $\Gamma^{\varepsilon,\theta}$ implies that the subgame perfect equilibrium $s^*$ of $\Gamma^{\varepsilon,\theta}$ prescribes an equilibrium of $G$ in the play phase when no agreements have been reached. □

In what follows, we fix an equilibrium $q^* = (q_1^*, \cdots, q_n^*)$ of $G$ given by an SSPE $s^*$ in the case of no agreements, and assume that $q^*$ satisfies Assumption 1.(iii). We denote the expected payoffs for $q^*$ by $d = (d_1, \cdots, d_n)$. We show that $d = (d_1, \cdots, d_n)$ becomes the disagreement point of the Nash bargaining solution when all players are active in negotiations. Note that the expected payoff $v = (v_1, \cdots, v_n)$ of $s^*$ satisfies $v_i \geq d_i$ for all $i \in N$ (if $v_i < d_i$ for some $i$, $i$ will obtain the expected payoff $(1-\varepsilon)v_i + \varepsilon d_i$ higher than $v_i$ by rejecting all proposals). Given that negotiation is meaningless if the disagreement point $d$ is Pareto-efficient, we assume:

**Assumption 2.** The disagreement point $d$ in an SSPE $s^*$ of $\Gamma^{\varepsilon,\theta}$ is Pareto-inefficient in the feasible set $F(G)$ of $G$, that is, there exists some $y = (y_1, \cdots, y_n) \in F(G)$ such that $y_i > d_i$ for all $i \in N$.

The following lemma proves Proposition 1, which shows no delay of agreement in every SSPE of $\Gamma^{\varepsilon,\theta}$.

**Lemma 2.** In every SSPE $s^* = (s_1^*, \cdots, s_n^*)$ of $\Gamma^{\varepsilon,\theta}$, every player's proposal is accepted in the initial round of the negotiation phase.

**Proof.** Let $v = (v_1, \cdots, v_n)$ be the expected payoffs of players for $s^*$, and let $F(G)$ be the feasible set of $G$. We note that $v \in F(G)$, because $F(G)$ is convex and $v$ is a convex combination of a finite number of points in $F(G)$. By Assumption 2, there exists $y = (y_1, \cdots, y_n)$ in $F(G)$ such that $y_i > d_i$ for all $i \in N$. Given that $y$ and $v$ are in the convex set $F(G)$, we have $(1-\varepsilon)v + \varepsilon y \in F(G)$ for any $\varepsilon$ with $0 < \varepsilon < 1$. Then, select $p^N \in \Delta(A^N)$ such that $u_j(p^N) = (1-\varepsilon)v_j + \varepsilon y_j$ for all $j \in N$. Given that $y_j > d_j$ for any $j$, we have

$$u_j(p^N) > (1-\varepsilon)v_j + \varepsilon d_j \quad \text{for all} \quad j \in N. \tag{5}$$

Suppose that every player $i$ proposes $(N, p^N)$. Given that $s^*$ is an SSPE of $\Gamma^{\varepsilon,\theta}$, the right-hand side of (5) is the expected payoff that player $j$ ($\neq i$) can obtain by rejecting the proposal $(N, p^N)$. (5) implies that every player $i$'s proposal $(N, p^N)$ is accepted by all other players. This fact implies that player $i$'s equilibrium proposal (not necessarily equal to $(N, p^N)$) is accepted on the equilibrium play of the SSPE $s^*$. $\square$

**Lemma 3.** Let $s^* = (s_1^*, \cdots, s_n^*)$ be an efficient SSPE of $\Gamma^{\varepsilon,\theta}$, $v = (v_1, \cdots, v_n)$ the expected payoffs of players for $s^*$, and $d = (d_1, \cdots, d_n)$ the disagreement payoff of $s^*$. In $s^*$, every player $i \in N$ initially proposes a pair $(N,\, p^i)$ where $p^i \in \Delta(A^N)$ is the optimal solution of the maximization problem

$$\begin{aligned} \max \quad & u_i(p) \\ \text{subject to} \quad & (1)\ p \in \Delta(A^N) \\ & (2)\ u_j(p) \geq (1-\varepsilon)v_j + \varepsilon d_j \quad \text{for all } j \in N,\ j \neq i. \end{aligned} \tag{6}$$

Moreover, the proposal $(N,\, p^i)$ is accepted.

**Proof.** Let $c_j^\varepsilon \equiv (1-\varepsilon)v_j + \varepsilon d_j$ denote the RHS of the second constraint in (6). If responder $j$ is offered more than $c_j^\varepsilon$, it is optimal for her to accept the proposal. (6) can be reformulated as

$$\begin{aligned} \max \quad & h_i(x_{-i}) \\ \text{subject to} \quad & (1)\ x_{-i} \in F_{-i}(G) \\ & (2)\ x_j \geq c_j^\varepsilon \quad \text{for all } j \in N,\ j \neq i \end{aligned}$$

Recall that $h_i(x_{-i}) = \max\{x_i \mid (x_i, x_{-i}) \in F(G)\}$. The function $h_i$ is continuous from Assumption 1(i). Let $x_{-i}^* \in R^{N-\{i\}}$ be the optimal solution of the problem above. It holds from Assumptions 1.(ii) and 1.(iii) that $x_j^* = c_j^\varepsilon$ for all $j \neq i$. For any $\varepsilon > 0$, $(c_j^\varepsilon)_{j\in N}$ is an interior point of the feasible set $F(G)$ (note that $v_j \geq d_j$ for all $j \in N$). Then, it holds from the continuity of $h_i$ that for sufficiently small $\delta_i > 0$, there exists $\delta_j > 0$ for all $j \neq i$ such that $h_i(x_{-i}^* + \delta_{-i}) \geq h_i(x_{-i}^*) - \delta_i$ where $\delta_{-i} = (\delta_j)_{j\neq i}$. This inequality means that if player $i$ proposes the grand coalition $N$ and the correlated action attaining payoffs $(x_{-i}^* + \delta_{-i}, h_i(x_{-i}^* + \delta_{-i}))$, this proposal is accepted and thus player $i$ can obtain more than $h_i(x_{-i}^*) - \delta_i$. Given that $\delta_i > 0$ can be chosen arbitrarily small, we can show that player $i$ proposes the optimal solution $p^i$ of (6) in the efficient SSPE $s^*$ of $\Gamma^{\varepsilon,\theta}$. Lemma 1 shows that the proposal is accepted. $\square$

Lemmas 2 and 3 characterize the equilibrium proposal of every player in an efficient SSPE of $\Gamma^{\varepsilon,\theta}$. We note that the optimal solution of the maximization problem in Lemma 3 gives only a necessary condition for the efficient SSPE proposal for every player $i$, because the optimality of proposing the grand coalition $N$ has not been examined. Given that player $i$ can propose any subcoalition $S$ of $N$, we must guarantee that the grand coalition $N$ is actually the optimal proposal. This will be done in Theorem 2 where the Nash core plays an important role. Before proving Theorem 2, we will prove that the maximization problem in Lemma 3 characterizes the asymmetric Nash bargaining solution of $G$ as the probability $\varepsilon$ that negotiations will fail goes to zero.

**Lemma 4.** Let $v = (v_1, \cdots, v_n)$ be a limit of efficient SSPE payoffs $v^\varepsilon = (v_1^\varepsilon, \cdots, v_n^\varepsilon)$ of $\Gamma^{\varepsilon,\theta}$ as $\varepsilon$ goes to zero. Then,

$$\frac{v_1 - d_1}{\theta_1} \cdot \frac{\partial H}{\partial x_1}(v) = \cdots = \frac{v_n - d_n}{\theta_n} \cdot \frac{\partial H}{\partial x_n}(v) \tag{7}$$

$$H(v) = 0 \tag{8}$$

where $\theta = (\theta_1, \cdots, \theta_n)$ is the probability distribution that selects a proposer from the player set $N$, and $d = (d_1, \cdots, d_n)$ is the disagreement payoff of an efficient SSPE in $\Gamma^{\varepsilon,\theta}$ (independent of $\varepsilon$).

**Proof.** Let $x_i^\varepsilon$ denote the payoff that every player $i \in N$ demands for herself in the initial round of the negotiation phase when the efficient SSPE of $\Gamma^{\varepsilon,\theta}$ is played. By Lemma 3, we can show that for every $i \in N$

$$H((1-\varepsilon)v_1^\varepsilon + \varepsilon d_1, \cdots, x_i^\varepsilon, \cdots, (1-\varepsilon)v_n^\varepsilon + \varepsilon d_n) = 0. \tag{9}$$

In addition, by Lemma 3 and the definition of $v^\varepsilon$, we obtain

$$v_i^\varepsilon = \theta_i x_i^\varepsilon + (1-\theta_i)[(1-\varepsilon)v_i^\varepsilon + \varepsilon d_i], \quad \text{for all} \quad i = 1, \cdots, n. \tag{10}$$

For each $i \in N$, define $z^{\varepsilon,i} \in F(G)$ as

$$z^{\varepsilon,i} = ((1-\varepsilon)v_1^\varepsilon + \varepsilon d_1, \cdots, x_i^\varepsilon, \cdots, (1-\varepsilon)v_n^\varepsilon + \varepsilon d_n). \tag{11}$$

$z^{\varepsilon,i}$ is the payoff vector proposed by player $i$ in the initial round of the negotiation phase in the efficient SSPE $s^\varepsilon$ of $\Gamma^{\varepsilon,\theta}$. For any $i,\ j \in N (i \neq j)$, we have from (9)

$$H(z^{\varepsilon,i}) - H(z^{\varepsilon,j}) = 0.$$

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

$$\begin{aligned} 0 &= H(z^{\varepsilon,i}) - H(z^{\varepsilon,j}) \\ &= [x_i^\varepsilon - (1-\varepsilon)v_i^\varepsilon - \varepsilon d_i] \cdot \frac{\partial H}{\partial x_i}(\lambda z^{\varepsilon,i} + (1-\lambda)z^{\varepsilon,j}) \\ &\quad + [(1-\varepsilon)v_j^\varepsilon + \varepsilon d_j - x_j^\varepsilon] \cdot \frac{\partial H}{\partial x_j}(\lambda z^{\varepsilon,i} + (1-\lambda)z^{\varepsilon,j}). \end{aligned} \tag{12}$$

(10) yields

$$x_i^\varepsilon - (1-\varepsilon)v_i^\varepsilon - \varepsilon d_i = \frac{1}{\theta_i}[v_i^\varepsilon - (1-\varepsilon)v_i^\varepsilon - \varepsilon d_i] = \frac{\varepsilon}{\theta_i}(v_i^\varepsilon - d_i). \tag{13}$$

By substituting (13) into (12), we prove

$$\frac{v_i^\varepsilon - d_i}{\theta_i}\frac{\partial H}{\partial x_i}(\lambda z^{\varepsilon,i} + (1-\lambda)z^{\varepsilon,j}) = \frac{v_j^\varepsilon - d_j}{\theta_j}\frac{\partial H}{\partial x_j}(\lambda z^{\varepsilon,i} + (1-\lambda)z^{\varepsilon,j}). \tag{14}$$

By assumption, we have $\lim_{\varepsilon\to 0} v^\varepsilon = v$, which implies from (10) that $\lim_{\varepsilon\to 0} x_i^\varepsilon = v_i$ for all $i$ . Thus, it follows from (11) that

$$\lim_{\varepsilon\to 0} v^\varepsilon = \lim_{\varepsilon\to 0} z^{\varepsilon,1} = \cdots = \lim_{\varepsilon\to 0} z^{\varepsilon,n} = v. \tag{15}$$

We prove (7) from (14) and (15), and prove (8) from (9) and (15). □

In view of (9) and (10), the efficient SSPE payoffs $v^{\varepsilon} = (v_1^{\varepsilon}, \cdots, v_n^{\varepsilon})$ of $\Gamma^{\varepsilon,\theta}$ are characterized as a solution of

$$v_i^{\varepsilon} = \theta_i \cdot h_i((1-\varepsilon)v_{-i}^{\varepsilon} + \varepsilon d_{-i}) + (1-\theta_i) \cdot \{(1-\varepsilon)v_i^{\varepsilon} + \varepsilon d_i\} \quad \text{for all } i \in N. \tag{16}$$

(16) is called the *equilibrium equation* of the efficient SSPE payoffs of $\Gamma^{\varepsilon,\theta}$.

We are now ready to prove Theorems 1 and 2.

**Proof of Theorem 1.** The maximization problem in Definition 2 is reformulated as

$$\begin{array}{ll} \max_x \sum_{i=1}^n & \theta_i \cdot \log(x_i - d_i) \\ \text{subject to} & (1)\ H(x_1, \cdots, x_n) \geq 0 \\ & (2)\ x_i \geq d_i \quad \text{for all} \quad i \in N. \end{array}$$

By Assumption 2, the optimal solution $x^* = (x_1^*, \cdots, x_n^*) \in R^N$ satisfies $H(x_1^*, \cdots, x_n^*) = 0$ and $x_i^* > d_i$ for all $i \in N$. Therefore, the Kuhn-Tucker condition yields

$$\begin{array}{l} \dfrac{\theta_i}{x_i^* - d_i} - \lambda \dfrac{\partial H}{\partial x_i}(x^*) = 0, \quad i = 1, \cdots, n \\ H(x^*) = 0 \end{array}$$

where $\lambda$ is the Lagrange multiplier. From the concavity of $H(x_1, \cdots, x_n)$ and Assumption 2, $x^*$ is the optimal solution of the maximization problem if and only if $x^*$ satisfies the Kuhn-Tucker condition. Together with this fact, Lemma 4 proves the theorem. □

**Proof of Theorem 2.** Let $\phi^*$ be the payoff configuration for $\Gamma^{\theta}$ that is generated by a limit-totally-efficient SSPE $s^* = (s_1^*, \cdots, s_n^*)$. Let $s^{\varepsilon} = (s_1^{\varepsilon}, \cdots, s_n^{\varepsilon})$

be a totally efficient SSPE of $\Gamma^{\varepsilon,\theta}$ that converges to $s^* = (s_1^*, \cdots, s_n^*)$ as $\varepsilon$ goes to zero. By the same proof as Lemma 1, we can show that for every correlated action $p_S \in \Delta(A^S)$ of every coalition $S$, $s^\varepsilon$ induces an equilibrium of the subgame $G(p_S)$ of $G$ when negotiations break down among all players in $N - S$. Let $d(p_S)$ denote the payoffs of such an equilibrium, and let $d$ denote the disagreement configuration of $G$ that assigns $d(p_S)$ to every subgame $G(p_S)$ of $G$. Let $\Gamma^{\varepsilon,\theta}(p_S)$ denote a subgame of $\Gamma^{\varepsilon,\theta}$ which starts after agreement $(S, p_S)$ has been reached. By applying Theorem 1 to every subgame $\Gamma^{\varepsilon,\theta}(p_S)$, we show that the payoff configuration $\phi^*$ satisfies (i).

We will next prove (ii). Let $x^* = (x_1^*, \cdots, x_n^*) \in R^N$ be the payoff vector that the payoff configuration $\phi^*$ assigns to the game $G$. For notational simplicity, we prove only that $x^*$ belongs to the Nash core of $G$ defined by the Nash bargaining solution configuration $b^*$ with $\theta$ and $d$. The same proof can be applied easily to the payoff vector $\phi^*(p_S)$, which the payoff configuration $\phi^*$ assigns to every correlated action $p_S$ of every $S$. Suppose that $x^*$ does not belong to the Nash core of $G$. By the definition of the Nash core, there exists some coalition $T \subset N$ and some payoff vector $y \in v^{Nash}(T)$ such that

$$y_i > x_i^* \quad \text{for all} \quad i \in T, \tag{17}$$

where $v^{Nash}$ is the Nash characteristic function (see Definition 5). By the definition of $v^{Nash}$, the fact that $y \in v^{Nash}(T)$ means that there exists some correlated action $p_T \in \Delta(A^T)$ of $T$ such that

$$u_i(p_T, b^*(p_T)) \geq y_i \quad \text{for all} \quad i \in T \tag{18}$$

where $b^*(p_T)$ is the Nash bargaining solution of the subgame $G(p_T)$. Let $b^\varepsilon(p_T)$ be the correlated action employed by the complementary coalition $N - T$ in the totally efficient SSPE $s^\varepsilon$ of $\Gamma^{\varepsilon,\theta}$ after $p_T$ is agreed by the coalition $T$. By

Theorem 1, we obtain

$$\lim_{\varepsilon \to 0} b^{\varepsilon}(p_T) = b^*(p_T). \tag{19}$$

Let $x^{\varepsilon} = (x_1^{\varepsilon}, \cdots, x_n^{\varepsilon})$ be the payoff vector of the totally efficient SSPE $s^{\varepsilon}$. Then,

$$\lim_{\varepsilon \to 0} x^{\varepsilon} = x^*. \tag{20}$$

In view of (17), (18), (19) and (20), it holds that for sufficiently small $\varepsilon > 0$

$$u_i(p_T, b^{\varepsilon}(p_T)) > x_i^{\varepsilon} \quad \text{for all} \quad i \in T. \tag{21}$$

Now, suppose that player $i \in T$ deviates from $s^{\varepsilon}$ and proposes $(T, p_T)$. If this proposal is agreed upon, then all responders $j$ in $T$ receive the payoff $u_j(p_T, b^{\varepsilon}(p_T))$, because thereafter the complementary coalition $N - T$ reacts to $T$ by choosing $b^{\varepsilon}(p_T)$. If the proposal $(T, p_T)$ is rejected, they receive the continuation payoff $(1-\varepsilon)x_j^{\varepsilon} + \varepsilon d_j$, which is smaller than $x_j^{\varepsilon}$ (note that $x_j^{\varepsilon} > d_j$). From (21), it is optimal for all responders in $T$ to accept $(T, p_T)$. Therefore, on the equilibrium play of $s^{\varepsilon}$, the proposal $(T, p_T)$ is agreed and the proposer $i$ is better-off. This contradicts the assumption that $s^{\varepsilon}$ is an SSPE of $\Gamma^{\varepsilon,\theta}$. $\square$

To prove Theorem 3, we first establish that there exists a solution for the equilibrium equation (16) of the efficient SSPE of $\Gamma^{\varepsilon,\theta}$.

**Lemma 5** Let $v = (v_1, \cdots, v_n)$ be the Nash bargaining payoffs of $G$ with weight vector $\theta = (\theta_1, \cdots, \theta_n)$ and disagreement point $d = (d_1, \cdots, d_n)$. For any sufficiently small $\varepsilon > 0$, there exists a solution $v^{\varepsilon} = (v_i^{\varepsilon})_{i \in N} \in F(G)$ to the equilibrium equation (16) such that $v^{\varepsilon}$ converges to $v$ as $\varepsilon$ goes to zero.

**Proof.** Let $F^* = \{x \in F(G) \mid x_i \geq d_i \text{ for all } i \in N\}$. For every $x \in F^*$

and every $i \in N$, define

$$g_i^\varepsilon(x) = \theta_i \cdot h_i((1-\varepsilon)x_{-i} + \varepsilon d_{-i}) + (1-\theta_i) \cdot \{(1-\varepsilon)x_i + \varepsilon d_i\}. \tag{22}$$

It can be proved that $g^\varepsilon(x) = (g_1^\varepsilon(x), \cdots, g_n^\varepsilon(x))$ is a continuous function from the compact convex subset $F^*$ of $R^n$ to itself. Then, by Brouwer's fixed point theorem, there exists a fixed point $v^\varepsilon \in F^*$ of $g^\varepsilon$ that satisfies (16). Given that $F^*$ is a compact set, there exists some converging subsequence of $\{v^\varepsilon\}$. Take any such subsequence of $\{v^\varepsilon\}$. Let $\bar{v}$ denote its limit. By the same proof as in Theorem 1 (and Lemma 4), we can prove $\bar{v} = v$. This implies that the sequence $\{v^\varepsilon\}$ itself has limit $v$. $\square$

Let $\phi^*$ be the Nash bargaining payoff configuration of $G$. By applying the same proof as Lemma 5 to every subgame $G(p_S)$ of $G$, we can show that there exists a solution to the equilibrium equation of an efficient SSPE of the subgame $\Gamma^{\varepsilon,\theta}(p_S)$ of $\Gamma^{\varepsilon,\theta}$. Let $v^\varepsilon(p_S)$ denote a solution. Lemma 5 also shows that $v^\varepsilon(p_S)$ converges to the Nash bargaining solution payoff $\phi^*(p_S)$ of $G(p_S)$ as $\varepsilon$ goes to zero.

**Proof of Theorem 3.** Let $d$ be a disagreement configuration of $G$ that satisfies (A). For every correlated action $p_S \in \Delta(A^S)$ of every coalition $S$, let $d(p_S) \in F(G(p_S))$ denote the disagreement point that the configuration $d$ assigns to the subgame $G(p_S)$ of $G$. With abuse of notation, we also denote by $d = (d_1, \cdots, d_n)$ the disagreement point in $G$ assigned by the disagreement configuration $d$.

Define every player $i$'s strategy $s_i^\varepsilon$ in $\Gamma^{\varepsilon,\theta}$ as follows.

(1) When no coalition forms,

(i) propose the grand coalition $N$ and the correlated action yielding the payoff vector in (11)

$$z^{\varepsilon,i} = (h_i((1-\varepsilon)v^{\varepsilon}_{-i} + \varepsilon d_{-i}), (1-\varepsilon)v^{\varepsilon}_{-i} + \varepsilon d_{-i}),$$

where $v^{\varepsilon} = (v^{\varepsilon}_i)_{i\in N} \in F(G)$ is a solution to the equilibrium equation (16) (the existence of which is proved in Lemma 5).

(ii) accept any proposal that yields a payoff not less than $(1-\varepsilon)v^{\varepsilon}_i + \varepsilon d_i$,

(iii) employ the Nash equilibrium of $G$ given by the disagreement configuration $d$ when negotiations break down.

(2) When some coalition $S$ forms and some correlated action $p_S \in \Delta(A^S)$ of $S$ is agreed upon, the strategy $s^{\varepsilon}_i$ is defined in the same way as above, except that $N$ and $v^{\varepsilon}$ are replaced with $N - S$ and $v^{\varepsilon}(p_S)$, respectively. When more than one coalition forms, $s^{\varepsilon}_i$ is defined in a similar way by taking $S$ as the union of coalitions.

Let $\phi^{\varepsilon}$ be the payoff configuration generated by the strategy profile $s^{\varepsilon} = (s^{\varepsilon}_1, \cdots, s^{\varepsilon}_n)$ constructed above. Given that $v^{\varepsilon}(p_S)$ is a solution to the equilibrium equation of an efficient SSPE of $\Gamma^{\varepsilon,\theta}(p_S)$ for every $p_S \in \Delta(A_S)$, we can show that $\phi^{\varepsilon}(p_S) = v^{\varepsilon}(p_S)$, and that $\phi^{\varepsilon}(p_S)$ converges to the Nash bargaining solution payoff $\phi^*(p_S)$ of $G(p_S)$ with $\theta$ and $d$ when $\varepsilon$ goes to zero.

It remains to be proved that the strategy profile $s^{\varepsilon} = (s^{\varepsilon}_1, \cdots, s^{\varepsilon}_n)$ is an SSPE of $\Gamma^{\varepsilon,\theta}(p_S)$. For this purpose, it is sufficient to prove that player $i$'s proposal $z^{\varepsilon,i}$ is optimal given $s^{\varepsilon}$. For each $j \in N$, let $z^{\varepsilon,i}_j$ denote the $j$-th component of player $i$'s proposal $z^{\varepsilon,i}$, that is,

$$z^{\varepsilon,i}_i = h_i((1-\varepsilon)v^{\varepsilon}_{-i} + \varepsilon d_{-i}), \quad z^{\varepsilon,i}_j = (1-\varepsilon)v^{\varepsilon}_j + \varepsilon d_j,\ j \neq i.$$

Given that the disagreement point $d = (d_1, \cdots, d_n)$ of $G$ is an interior point of $F(G)$ from Assumption 2, $(1-\varepsilon)v^{\varepsilon} + \varepsilon d$ is also an interior point of $F(G)$ (note that $F(G)$ is a convex set of $R^N$). This implies that $h_i((1-\varepsilon)v^{\varepsilon}_{-i} + \varepsilon d_{-i}) > (1-\varepsilon)v^{\varepsilon}_i + \varepsilon d_i$ for every $i \in N$. Then, it follows from (16) that $z^{\varepsilon,j}_i < v^{\varepsilon}_i < z^{\varepsilon,i}_i$ for any $j \neq i$. In addition, we can see from Lemma 5 that $v^{\varepsilon}$ and every $z^{\varepsilon,i}$ converge to the Nash bargaining solution payoffs $v$ of $G$ with the weights $\theta = (\theta_1, \cdots, \theta_n)$ and the disagreement point $d = (d_1, \cdots, d_n)$ as $\varepsilon$ goes to zero. Given that $v$ belongs to the interior (relative to the upper-right boundary $H$ of the feasible set $F(G)$) of the strict Nash core of $G$ and $z^{\varepsilon,i}$ belongs to the boundary $H$, we can see that $z^{\varepsilon,i}$ also belongs to the (relative) interior of the strict Nash core for any sufficiently small $\varepsilon$. Take any coalition $S$ and any correlated action $p_S$ of $S$. By definition, the payoff vector $u = (u_j(p_S, b^*(p_S)))_{j \in N}$ is Nash-effective for $S$, that is, $u \in v^{Nash}(S)$. The fact that $z^{\varepsilon,i}$ is in the strict Nash core implies that if $u_j(p_S, b^*(p_S)) \geq z_j^{\varepsilon,i}$ for all $j \in S, j \neq i$, then $z_i^{\varepsilon,i} \geq u_i(p_S, b^*(p_S))$. Otherwise, $u$ dominates $z^{\varepsilon,i}$ via $S$ in the strict sense with respect to $v^{Nash}$. Therefore, $z_i^{\varepsilon,i}$ is the optimal value (attained by $S = N$) of the maximization problem

$$\begin{aligned} &\max \ u_i(p_S, b^*(p_S)) \\ &\qquad \text{subject to} \qquad (1)\ S \subset N,\ p_S \in \Delta(A^S) \\ &\qquad\qquad\qquad\qquad\quad (2)\ u_j(p_S, b^*(p_S)) \geq z_j^{\varepsilon,i} \quad \text{for all} \quad j \in S,\ j \neq i. \end{aligned}$$

This means that the strategy $s_i^{\varepsilon}$ prescribes the optimal proposal of player $i$. It is clear that $s_i^{\varepsilon}$ prescribes the optimal action for responders. By applying the same proof to all subgames of $\Gamma^{\varepsilon,\theta}$ starting after some agreement has been reached, we can prove that $s^{\varepsilon} = (s_1^{\varepsilon}, \cdots, s_n^{\varepsilon})$ is a totally efficient SSPE of $\Gamma^{\varepsilon,\theta}$. □

# 6 Discussion

We discuss several issues that our noncooperative approach may raise. First, how does the noncooperative model presented herein yield the Nash core? The following properties of our model are crucial for any bargaining game to yield the Nash core: (i) the Nash bargaining solution is implemented on the equilibrium path, (ii) any coalition of players is free to deviate from the largest coalition, and (iii) the bargaining game has a recursive structure in the sense that if any coalition deviates from the equilibrium agreement, then other players play the same kind of a bargaining game among themselves as the original one. As we argued in the introduction, subgame perfection and properties (i) and (iii) imply that the other players employ the Nash bargaining solution behavior in response to any deviation by a coalition. Then, property (ii) yields that the Nash bargaining solution of the largest coalition belongs to the Nash core. By replacing the Nash bargaining solution with the core, the same arguments as above can be applied to noncooperative implementation of the recursive core that are presented by Huang and Sjöström [14] and Kóczy [17]. Indeed, by employing the bargaining model of Perry and Reny [26] for the core, Huang and Sjöström [15] provide noncooperative implementation of the recursive core for partition function form games. Kóczy [16] obtains the same result by employing the bargaining model of Moldovanu and Winter [19].

Second, both our model and Huang and Sjöström's [15] model assume that members of any coalition can commit to their agreement. There, however, is a slight difference between the two models regarding a form of agreements. In Huang and Sjöström's [15] framework of a partition function form game, the members of a coalition need to commit to a division rule (or, in general, a complete contingency plan regarding how to divide a coalitional surplus) for every possible coalition structure because the value of a coalition is not well-defined until a final coalition structure is determined. In our framework of a strategic form game, given that the members of a coalition can anticipate rationally that other players will form the complementary coalition, it is sufficient for them to commit to a profile of actions against it.

Third, the noncooperative bargaining theory is often criticized on the ground that the result is too sensitive to unimportant details of bargaining procedures. For example, different orders of proposers may produce different bargaining outcomes even if the fundamental parameters of the worth of a coalition remain unchanged. In our model, the equilibrium outcome depends on the probability distribution for selecting a proposer in each bargaining round. In our view, this random rule should not be regarded as an unimportant procedural detail. The probability distribution turns out to induce weights of players in the generalized Nash bargaining solution. The result supports nicely our intuition that the likelihood that a player will become a proposer is a source of her bargaining power.

Fourth, the probability distribution for selecting a proposer is predetermined in our bargaining model. A natural question is how such a probability distribution is determined in a real situation. This is an empirical matter that lies beyond the scope of this paper. Social and political factors may determine a probability distribution. For example, in local communities, a seniority rule (older persons propose more often than younger ones) tends to prevail. In international negotiations, "bigger" countries with large populations (or GDPs) may be given more opportunities to make proposals than others.

Fifth, our analysis focuses on an efficient equilibrium in which the grand coalition forms. We have proved that given a disagreement point, a (limit) totally efficient SSPE exists uniquely if the Nash bargaining solution belongs to the Nash core. The uniqueness of an SSPE is an open question in the random proposer model for a general cooperative game. In a special class of TU games without externality, Yan [34] proves that an efficient SSPE is a unique SSPE when it belongs to the core. Eraslan [8] proves the uniqueness of an SSPE in an $n$-person majority game. There is a further issue regarding the uniqueness of an equilibrium in our bargaining model. The disagreement point of the Nash bargaining solution is determined by a Nash equilibrium of a strategic form game, which is a primitive of the analysis. Obviously, in order to derive a unique SSPE of the model, we need an equilibrium selection theory for a strategic form game.

Finally, note there is a noticeable difference between our model and Huang and Sjöström's [15] model when core solutions are empty. When the recursive core is empty, Huang and Sjöström's model has no SSPE (in pure strategies). When the Nash core is empty, our model does not have the (limit) totally efficient SSPE, but another type of SSPE exists.[18] For example, a majority voting game has Baron-Ferejohn's [2] equilibrium, in which all minimal winning coalitions have the same probability of forming. In general, an SSPE of the random proposer model is inefficient when the Nash core is empty. When the resulting payoff allocation is inefficient, players may want to renegotiate their on-going agreement to attain a Pareto-improving payoff allocation. This problem of renegotiations in coalitional bargaining is analyzed by Okada [24] and Seidmann and Winter [30] in a TU game in characteristic function form. Gomes [9] and Gomes and Jehiel [10] extend the renegotiation analysis to a general bargaining situation with externality. All these studies show that successive renegotiations necessarily lead to an efficient allocation when the grand coalition is efficient. We, however, show in Okada [24] a negative effect of renegotiations such that they may distort the equity of a final allocation by inducing the first-mover rent.

# 7 Conclusion

We have extended a noncooperative theory of the Nash bargaining solution to a general $n$-person cooperative game. When a coalition forms, the reaction by other players is crucial for determining a final outcome of the game. The noncooperative approach yields a new core notion called the Nash core, which requires that the complementary coalition should react according to the Nash bargaining solution. We have proved that the Nash bargaining solution needs to belong to the Nash core of the game in order to be supported by a noncooperative equilibrium of the bargaining model.

[18]The existence of an SSPE in our model can be proved in a standard way by Kakutani's fixed-point theorem with induction on the number of players. See Ray and Vohra [27] and Gomes [9] for the proof in related models.

# References

[1] R.J. Aumann, The core of a cooperative game without side payments, Trans. Amer. Math. Soc. 98 (1961) 539-552.

[2] D. P. Baron, J. A. Ferejohn, Bargaining in legislatures, Amer. Polit. Sci. Rev. 83 (1989) 1181-1206.

[3] D. Baron, E. Kalai, The simplest equilibrium of a majority-rule division game, J. Econ. Theory 61 (1993) 290-301.

[4] K. G. Binmore, A. Rubinstein, A. Wolinsky, The Nash bargaining solution in economic modelling, RAND J. Econ. 17 (1986) 176-188.

[5] S. K. Chakrabarti, Refinements of the $\beta$-core and the strong equilibrium and the Aumann proposition, Int. J. Game Theory 17 (1988) 205-224.

[6] P. Chander, H. Tulkens, The core of an economy with multilateral environmental externalities, Int. J. Game Theory 26 (1997) 379-401.

[7] K. Chatterjee, B. Dutta, D. Ray, K. Sengupta, A noncooperative theory of coalitional bargaining, Rev. Econ. Stud. 60 (1993) 463-477.

[8] H. Eraslan, Uniqueness of stationary equilibrium payoffs in the Baron-Ferejohn model, J. Econ. Theory 103 (2002) 11-30.

[9] A. Gomes, Multilateral contracting with externalities, Econometrica 73 (2005) 1329-1350.

[10] A. Gomes, P. Jehiel, Dynamic processes of social and economic interactions: On the persistence of inefficiencies, J. Polit. Economy 113 (2005) 626-667.

[11] J.C. Harsanyi, A simplified bargaining model for the $n$-person cooperative game, Int. Econ. Rev. 4, (1963) 194-220.

[12] J.C. Harsanyi, R. Selten, A General Theory of Equilibrium Selection in Games, The MIT Press, Cambridge, MA, 1988.

[13] S. Hart, An axiomatization of Harsanyi's non-transferable utility solution, Econometrica 53 (1985) 1295-1313.

[14] C.-Y. Huang, T. Sjöström, Consistent solutions for cooperative games with externalities, Games Econ. Behav. 43 (2003) 196-213.

[15] C.-Y. Huang, T. Sjöström, Implementation of the recursive core for partition function form games, J. Math. Econ. 42 (2006) 771-793.

[16] L. Á. Kóczy, Sequential coalition formation and the core in the presence of externalities, METEOR Research Memorandum 06/047, Maastricht University, 2006.

[17] L. Á. Kóczy, A recursive core for partition function form games, Theory Dec. 63 (2007) 41-51.

[18] E. Maskin, J. Tirole, Markov perfect equilibrium: I. Observable actions, J. Econ. Theory 100 (2001) 191-219.

[19] B. Moldovanu, E. Winter, Order independent equilibria, Games Econ. Behav. 9 1995 21-34.

[20] M. Montero, Endogenous coalition formation and bargaining, Dissertation Series 71, Center for Economic Research, Tilburg University, 2000.

[21] J.F. Nash, The bargaining problem, Econometrica 18 (1950) 155-162.

[22] J.F. Nash, Two-person cooperative games, Econometrica 21 (1953) 128-140.

[23] A. Okada, A noncooperative coalitional bargaining game with random proposers, Games Econ. Behav. 16 (1996) 97-108.

[24] A. Okada, The efficiency principle in non-cooperative coalitional bargaining, Japanese Econ. Rev. 51 (2000) 34-50.

[25] M. J. Osborne, A. Rubinstein, Bargaining and Markets, Academic Press, New York, NY, 1990.

[26] M. Perry, P.J. Reny, A noncooperative view of coalition formation and the core, Econometrica 62 (1994) 795-817.

[27] D. Ray, R. Vohra, A theory of endogenous coalition structures, Games Econ. Behav. 26 (1999) 286-336.

[28] A. Rubinstein, Perfect equilibrium in a bargaining model, Econometrica 50 (1982) 97-109.

[29] H.E. Scarf, On the existence of a cooperative solution for a general class of $n$-person games, J. Econ. Theory 3 (1971) 169-181.

[30] D.J. Seidmann, E. Winter, A theory of gradual coalition formation, Rev. Econ. Stud. 65 (1998) 793-815.

[31] R. Selten, A noncooperative model of characteristic-function bargaining, in: V. Böhm, H. Nachtkamp (Eds.), Essays in Game Theory and Mathematical Economics in Honor of Oscar Morgenstern, Bibliographisches Institut, Mannheim, 1981, pp.131-151.

[32] P. P. Shenoy, On coalition formation: A game-theoretical approach, Int. J. Game Theory 8 (1979) 133-164.

[33] J. von Neumann, O. Morgenstern, Theory of Games and Economic Behavior, Princeton University Press, Princeton, NJ, 1944.

[34] H. Yan, Noncooperative selection of the core, Int. J. Game Theory 31 (2002) 527-540.