The Japanese Economic Review
The Journal of the Japanese Economic Association

The Japanese Economic Review doi: 10.1111/jere.12058
Vol. 66, No. 1, March 2015

COOPERATION AND INSTITUTION IN GAMES

By AKIRA OKADA
Hitotsubashi University

Based on recent developments in non-cooperative coalitional bargaining theory, I review game theoretical analyses of cooperation and institution. First, I present basic results of the random-proposer model and apply them to the problem of involuntary unemployment in a labour market. I discuss extensions to cooperative games with externalities and incomplete information. Next, I consider the enforceability of an agreement as an institutional foundation of cooperation. I re-examine the contractarian approach to the problem of cooperation from the viewpoint that individuals may voluntarily create an enforcement institution.
JEL Classification Numbers: C71, C72, C78, D02.

1. Introduction

Since the foundation provided by von Neumann and Morgenstern (1944), game theory has developed as a mathematical theory to investigate economic behaviour involving conflict and cooperation. In the global society, we are faced with many economic, political and social problems. These include monetary crises, unemployment, international trade, territorial conflicts, natural resources and the environment. Therefore, it is more important than ever that we scrutinize, theoretically and practically, whether and how we (as players) can cooperate and resolve these conflicts. In this paper, I review recent game theoretical analyses on cooperation and institutions.

There are numerous mechanisms by which cooperation is sustained among individual players who pursue their own goals. These include kin, evolution, reciprocity, altruism, trust, communication, learning, reputation, social norms, negotiations and institutions.[1] These mechanisms should work in a complementary fashion to promote cooperation and, in general, social order. My exposition focuses on negotiation and institutions, which play important roles in a society as it grows beyond the primitive stage.

The first part of the paper reviews recent developments on non-cooperative coalitional bargaining theory. Since the work of von Neumann and Morgenstern (1944), various solutions to the coalitional bargaining problem have been proposed in cooperative game theory. While many solutions are based on innovative ideas on group behaviour, there has been no consensus among game theorists on an appropriate solution for an *n*-person cooperative game. It may be argued that the diversity of solutions is a virtue, reflecting the complexity of the real world. However, to apply game theory to economic analyses, we need a general framework to understand when one solution is more suitable than others. Cooperative solution theory for economic situations that include externalities and incomplete information has not been well explored.

[1] While competition is often emphasized as the primary function of market mechanisms, Adam Smith (1776) considered the roles of division of labour and cooperation by economic agents. Competition may be regarded as an element of an overall process of negotiation in a market.

The non-cooperative game approach to the problem of cooperation was initiated in the seminal works of Nash (1951, 1953), and is called the Nash Program. The approach aims to explain cooperation as the result of individual players' payoff maximization in an equilibrium of a non-cooperative bargaining game that models pre-play negotiations.[2] Cooperation should be strategically stable. The non-cooperative approach is suitable for studying how the outcomes of economic activity are determined by negotiation rules, belief and strategic incentives. The approach re-examines a widely held view in economics, called the efficiency principle, that a Pareto-efficient allocation of resources can be attained through voluntary bargaining by rational agents if there is neither private information nor bargaining costs. The principle has been argued as the primary part of the celebrated Coase Theorem (see Coase, 1960; Cooter, 1989).

The second part of the paper considers institutional foundations for cooperation. Institutional arrangements facilitate cooperation in a society. The enforceability of agreements is one of the most critical conditions for cooperation. In most bargaining models, it is assumed that an agreement of cooperation can be enforced once it is reached among bargainers.[3] How can an agreement of cooperation be enforced?

In this paper, I refer to a social mechanism to enforce an agreement as an *institution*. Institutions take on diverse forms. Some institutions, such as the police and courts, are centralized in the sense that a central authority sanctions violators. Others are decentralized, and mutual monitoring and punishments among agents prevent agreements from being violated. Examples of decentralized institutions are social norms, convention and community enforcement, which are often formalized as a repeated game equilibrium.

Here, I address the question of how an institution emerges in a society by re-examining the contractarian point of view that individuals may voluntarily agree to create an institution for their collective benefit. There is a well-known puzzle in the institutional approach to cooperation: because rational individuals with self-interest have an incentive to free ride on an institution that enhances cooperation, they are likely to fail in forming institutions.[4] I review recent works on institution formation in a social dilemma situation in which the pursuit of individual interests conflicts with maximizing social welfare. Classic examples of this social dilemma include public goods provision and common-pool resource management.

The paper is organized as follows. Section 2 reviews recent works on non-cooperative coalitional bargaining theory. Here, I present the basic results of the random-proposer model. Then, I apply the theory to the issue of involuntary unemployment in a labour market, and discuss extensions to cooperative games with externalities and incomplete information. Section 3 reviews recent work on institution formation in social dilemma situations. Finally, Section 4 concludes the paper.

---

[2] Nash (1951, p. 295) explains his approach as follows: "One proceeds by constructing a model of the pre-play negotiation so that the steps of negotiation become moves in a larger non-cooperative game [. . .] describing the total situation. . . . Thus, the problem of analysing a cooperative game becomes the problem of obtaining a suitable, and convincing, non-cooperative model for the negotiation".

[3] If any agreement cannot be enforced, then a bargaining game is simply a cheap-talk game in which a "babbling equilibrium" without cooperation always exists.

[4] Kosfeld *et al*. (2009) term this puzzle a "dilemma of endogenous institution formation". The dilemma is sometimes called the "second-order free rider problem" (Oliver, 1980).

2. Theory of cooperation

2.1 Non-cooperative bargaining theory of coalition formation: An overview

I start by briefly reviewing current literature on the non-cooperative *n*-person bargaining theory of coalition formation.[5] Following Nash's (1953) pioneering study on the two-person bargaining game, Harsanyi (1974) presents a non-cooperative bargaining model, in extensive form, for an *n*-person game in characteristic function form. This model interprets the von Neumann and Morgenstern (1944) solution (i.e. stable set) as an equilibrium point of the bargaining game. Then, Selten (1981) presents a sequential bargaining game in which players propose coalitions and feasible payoff allocations until an agreement is reached. Selten shows that an equilibrium in his model is closely connected to the cooperative solution called a stable demand vector (Albers, 1975).

Since the seminal work of Selten (1981), the literature on non-cooperative coalitional bargaining has received widespread interest from researchers, and is now actively growing. While most works attempt to reconstruct cooperative solutions as equilibrium outcomes of non-cooperative sequential bargaining games, in the spirit of the Nash Program, they are motivated by two different (but closely related) research interests.

The first line of research is the non-cooperative foundation of cooperative solutions. This research has been carried out from both normative and positive perspectives. A typical problem from a normative point of view can be stated as follows: how can one (as a rule maker) design a well-defined bargaining procedure that implements some cooperative solution as an equilibrium outcome?[6] From a positive point of view, a typical question is as follows: can a cooperative solution be sustained as an equilibrium point of a non-cooperative bargaining game that suitably describes a negotiation process in the real world? If the answer to the question is negative, then the cooperative solution is no longer relevant.

Major solution concepts in cooperative game theory have been studied using the non-cooperative equilibrium approach, as well as using the cooperative axiomatic approach. The core and the Shapley value are the most studied solution concepts. Non-cooperative bargaining models for the core have been proposed in several works, including Okada (1992), Perry and Reny (1994), Moldovanu and Winter (1994, 1995), Okada and Winter (2002), Serrano (1995), Serrano and Vohra (1997), Evans (1997) and Horniaček (2008), among others. Non-cooperative bargaining models for the Shapley value are introduced by Gul (1989), Hart and Mas-Colell (1996) and Pérez-Castrillo and Wettstein (2001).

The second line of research aims to establish a positive theory of coalitional bargaining to understand how economic agents behave in multilateral negotiations, and what outcomes prevail in coalition formation and payoff allocation. Specifically, in the framework of non-cooperative coalitional bargaining, the literature has re-examined the efficiency principle. Chatterjee *et al*. (1993) extend the Rubinstein (1982) alternating-offers model to

---

[5] Because this overview is to provide readers with the necessary context for the expositions in the paper, some important contributions to the literature are not included. Bandyopadhyay and Chatterjee (2006), Ray (2007), and Ray and Vohra (2014) provide excellent overall summaries of the current literature.

[6] This problem is closely related to a more general framework of mechanism design or implementation, introduced by Hurwicz (1960): how can one design a mechanism that implements some social goal as an equilibrium outcome. Since Nash (1953), the literature on non-cooperative bargaining theory has attempted to obtain a suitable and convincing model (mechanism) for negotiations in the absence of a social planner.

coalitional bargaining. In their model, the first proposer is determined by a fixed order over players, and the first rejector becomes the next proposer. Proposals and responses are repeated until all players join (possibly different) coalitions. Players discount their future payoffs. Chatterjee *et al*. (1993) show that an agreement may be delayed in a stationary subgame perfect equilibrium (SSPE) of their "rejector-proposes" model,[7] and that the efficiency principle does not necessarily hold owing to the formation of an inefficient subcoalition when players are sufficiently patient. Players may not agree to form a grand coalition in an SSPE, even if it is a unique Pareto efficient coalition. Ray and Vohra (1999) extend the rejector-proposes model to a game with widespread externalities in partition function form, where the value of a coalition depends on the entire coalition structure.

Baron and Ferejohn (1989) propose another generalization of the two-person Rubinstein-type sequential bargaining game for legislative bargaining, described as an n-person simple majority game. At the beginning of every round, one player is randomly selected as a proposer according to a uniform probability distribution. The selected player proposes a winning coalition and a payoff allocation of coalition members. If the proposal is rejected by any member, then the next round is repeated using the same rule. The game continues until a winning coalition forms. Baron and Ferejohn prove the existence of an SSPE and the uniqueness of an SSPE payoff.

Legislative bargaining as a formal process is conducted according to a concrete rule specifying who may make proposals and how the proposals are agreed. A non-cooperative coalitional bargaining game is well suited to the analysis of legislative bargaining. The Baron and Ferejohn model characterizes a voting equilibrium reflecting the structures of legislatures in which the procedure-free model of social choice theory (or the core theory) yields no equilibrium. Since the seminal work of Baron and Ferejohn, their "random-proposer" model has been studied intensively, both theoretically and empirically, in the field of legislative bargaining. For further information, see Banks and Duggan (2000), Eraslan (2002), Snyder *et al*. (2005) and Adachi and Watanabe (2007).

Okada (1996) considers the random-proposer model of coalition formation in an n-person super-additive game in characteristic function form, and proves that no delay of agreement can occur in an SSPE of the model, unlike the rejector-proposes model. The reason for this difference in the two bargaining models is that if a responder rejects a proposal, then he or she runs the risk of not being selected as the next proposer in the random-proposer model and, thus, being excluded from a profitable coalition in future negotiations. As a result, all responders' continuation payoffs, being equal to their acceptance thresholds, may be smaller in the random-proposer model than in the rejector-proposes model. Owing to the decrease in responders' bargaining power, a proposer can make an optimal and acceptable proposal in the random-proposer model. It is also proved that, when players are sufficiently patient, a grand coalition is formed with an equal allocation (regardless of who becomes a proposer) if and only if the grand coalition has the largest coalitional value per capita. This condition is equivalent to the condition that the equal allocation belongs to the core of the underlying cooperative game. As Chatterjee *et al*. (1993) show, the result holds true in their rejector-proposes model, independent of an initial proposer. Thus, the aforementioned property of efficiency and equity in coalitional

---

[7] It is well known that all individually rational payoff allocations can be supported as (history-dependent) subgame perfect equilibria for high discount factors in Rubinstein-type sequential bargaining games with more than two players, even when no coalition is allowed. See Sutton (1986) and Osborne and Rubinstein (1990).

bargaining is robust with respect to changes in rules that govern the selection of proposers. The random-proposer model of coalitional bargaining has been studied extensively by, among others, Okada (2000, 2010, 2011), Yan (2002), Montero (2002, 2006), Gomes (2005), Hyndman and Ray (2007), Laruelle and Valenciano (2008), Kawamori (2008), Miyakawa (2009) and Compte and Jehiel (2010).

2.2 The model

An *n-person game in coalitional form with transferable utility* is represented by a pair (N, v). Here, $N = \{1, 2, \cdots, n\}$ is the set of players. A non-empty subset S of N (including $S = N$) is called a *coalition* of players. Let $C(N)$ be the set of all coalitions of N. The *characteristic function* v is a real-valued function on $C(N)$ satisfying: (i) (zero-normalized) $v(\{i\}) = 0$, for all $i \in N$; (ii) (super-additive) $v(S \cup T) \geq v(S) + v(T)$, for any two disjoint coalitions S and T; and (iii) (essential) $v(N) > 0$. For each S, $v(S)$ is interpreted as a sum of money that the members of S can distribute among themselves in any way if they agree to a payoff distribution. The cardinality of S is denoted by $|S|$.

A payoff allocation for coalition S is a vector $x^S = (x_i^S)_{i \in S}$ of real numbers, where x_i^S represents a payoff for player $i \in S$. A payoff allocation x^S for S is *feasible* if $\sum_{i \in S} x_i^S \leq v(S)$. Let X^S denote the set of all feasible payoff allocations for S, and let X_+^S denote the set of all elements in X^S with non-negative components. For a finite set Y, let $\Delta(Y)$ denote the set of all probability distributions on Y.

Let p be a function that assigns to every coalition $S \in C(N)$ a probability distribution $p^S \in \Delta(S)$. I refer to p as the *recognition probability*.

The *random-proposer model* represents a non-cooperative bargaining procedure for a game (N, v) as follows. Negotiations in coalition formation and payoff allocation take place over a (possibly) infinite number of rounds, t (= 1, 2, $\cdots$). Once players agree to form a coalition, they exit the game. Let $N^t (\subset N)$ be the set of all players who remain in the game in round t. Initially, I set $N^1 = N$. At the start of each round, one player, $i \in N^t$, is selected as a proposer according to the probability distribution $p^{N^t} \in \Delta(N^t)$. The recognition probability p is given exogenously. Player i proposes coalition S, with $i \in S \subset N^t$, and a payoff allocation, $x^S \in X_+^S$. All other members in S either accept or reject the proposal (S, x^S) sequentially. The order of responders does not affect the result in any critical way. If all responders accept the proposal, then coalition S forms and all its members exit the game. Thereafter, negotiations proceed to the next round, $t + 1$, and the process is repeated with $N^{t+1} = N^t - S$. Otherwise, negotiations continue in the next round with $N^{t+1} = N^t$. The game ends when no players remain in the negotiations.

The payoffs of players are defined as follows. When a proposal (S, x^S) is agreed in round t, every player $i \in S$ receives $\delta_i^{t-1} x_i^S$, where δ_i $(0 \leq \delta_i < 1)$ is the discount factor for future payoffs for player i. When the game does not stop, all remaining players receive zero payoffs. All players have perfect information about the history of the play whenever they choose an action. The above bargaining game is denoted by $\Gamma(N, p, \delta)$, where $\delta = (\delta_1, \cdots, \delta_n)$.

Interpretation: The random selection of a proposer in the bargaining model may be interpreted in several ways. First, the model can be interpreted so that the random choice of a proposer is actually employed as a formal rule in negotiations. Because a proposer may have an advantage in agreement, all players want to be selected as a proposer. As a tie-breaking rule, the random device seems to be a natural rule to select a proposer.

Second, an alternative interpretation is that the model describes a bargaining situation in which we, as analysts, observe that all or some players have opportunities to propose with different or equal likelihoods. Even if the analyst cannot observe a real process that determines a proposer, the model can give us an appropriate description of the process consistent with such an empirical observation. For example, in many multi-party parliament systems, the party with the largest number of seats tends to be recognized as most likely to form a government. The random-proposer model has been extensively applied to the study of government formation in the field of legislative bargaining. Third, there may be many kinds of random events in which the outcomes critically affect negotiations in economic situations. Random encounters in labour markets are such examples. Even if workers have the same skills, some workers may be employed and others may not, owing to a random event of encounters. The random selection of a proposer is a way to formulate the randomness and strategic behaviour in coalition formation. Finally, a critical factor of the random-proposer model is that the first rejector does not necessarily have an opportunity to make a counter-proposal, unlike in the rejector-proposes model. The rejector runs the risk of not being selected as a proposer and, thus, not joining future coalitions. Such a risk plays a critical role in players' responses. We may interpret the recognition probability of a player as his or her subjective estimate about the risk of being left out of future negotiations.

A *(behaviour) strategy*, denoted by σ_i, for player i in $\Gamma(N, p, \delta)$ is defined in a standard manner. Roughly, the strategy assigns the player's (random) action to his or her every move, depending on the history of game play. For a strategy combination $\sigma = (\sigma_1, \cdots, \sigma_n)$ of players, the expected (discounted) payoff for player i in $\Gamma(N, p, \delta)$ is defined in the usual way.

Definition 1: *A strategy combination $\sigma = (\sigma_1, \cdots, \sigma_n)$ of $\Gamma(N, p, \delta)$ is called a* stationary subgame perfect equilibrium (SSPE) *if σ is a subgame perfect equilibrium of $\Gamma(N, p, \delta)$ and the strategy σ_i of every player i depends only on the payoff-relevant history that consists of the player set N^t in every round t.*[8]

To present the fundamental results of the random-proposer model, I introduce the following notation. For SSPE σ of $\Gamma(N, p, \delta)$ and each $S \subset N$, let v_i^S denote the expected payoff for player i in the random-proposer model $\Gamma(S, p, \delta)$, where the player set is restricted to S. Then, let $q_i^S \in \Delta(\{T \mid i \in T \subset S\})$ denote the player's random choice of coalitions, T of S, including the player. I refer to a collection $(v^S, q^S)_{S \in C(N)}$, with $v^S = (v_i^S)_{i \in S}$ and $q^S = (q_i^S)_{i \in S}$, as the *configuration* of σ.

Theorem 1. *(Okada, 1996*, 2011*)*

(i) There exists an SSPE in the bargaining game $\Gamma(N, p, \delta)$ for every p and every δ.
(ii) For every SSPE σ of $\Gamma(N, p, \delta)$, every proposal is accepted in the initial round. All responders j are offered their discounted expected payoffs, $\delta_j v_j^N$. Then, $v_i^N > 0$ holds for every $i \in N$ with $p_i > 0$.

---

[8] When player i is a responder, the history includes the current proposal.

(iii) A collection $(v^S, q^S)_{S\in C(N)}$, with $v^S = (v_i^S)_{i\in S}$ and $q^S = (q_i^S)_{i\in S}$, is the configuration of an SSPE in $\Gamma(N, p, \delta)$ if and only if the following conditions hold for every $S \in C(N)$ and every $i \in S$:

(a) If q_i^S chooses coalition $\hat{S}$ with a positive probability, then $\hat{S}$ is a solution of

$$\max_{i\in T\subset S}\left(v(T) - \sum_{j\in T, j\neq i} \delta_j v_j^S \right). \tag{1}$$

(b) $v_i^S \in R_+$ satisfies

$$\begin{aligned} v_i^S = {} & p_i^S \max_{i\in T\subset S}\left(v(T) - \sum_{j\in T, j\neq i} \delta_j v_j^S \right) \\ & + \sum_{j\in S, j\neq i} p_j^S \delta_i \left(\sum_{j\in T\subset S, i\in T} q_j^S(T) v_i^S + \sum_{j\in T\subset S, i\notin T} q_j^S(T) v_i^{S-T} \right). \end{aligned} \tag{2}$$

Theorem 1 shows that an SSPE always exists in behaviour strategy. An SSPE in pure strategy does not always exist. By the stationarity of an SSPE, every responder j receives his or her discounted expected payoff, $\delta_j v_j^S$, where S is the player set in the game if he or she rejects a proposal. Thus, the value $\delta_j v_j^S$ becomes the player's acceptance level. Therefore, proposer i receives the "residual payoff", $v(T) - \sum_{j\in T, j\neq i} \delta_j v_j^S$, if he or she proposes coalition T, offering exactly $\delta_j v_j^S$ to all members j of T. Equation (1), referred to as the *optimality condition*, shows that the proposer maximizes his or her residual payoff in the selection of a coalition. In Equation (2), the expected payoff v_i^S of each player i consists of two parts, according to the rule of the random-proposer model. The first term on the right-hand side of Equation (2) shows player i's residual payoff when he or she is selected as a proposer. The second term shows the player's payoffs when he becomes a responder. Two possible cases should be considered. If player i is invited to join a coalition $T \subset S$, then he or she can receive the acceptance payoff $\delta_i v_i^S$. If not, the game proceeds to the next round, and he or she will receive the discounted expected payoff, $\delta_i v_i^{S-T}$. Equation (2) is referred to as the *payoff equation*. These two conditions fully characterize an SSPE for every player set $S \subset N$, given the supports of all players' random choices, q_i^S (i.e. the set of all coalitions, T, to which q_i^S assigns a positive probability).

A main issue in non-cooperative bargaining theory is under what condition an efficient allocation of payoffs can be voluntarily agreed by rational individuals. I consider the efficiency problem in the random-proposer model with the help of Theorem 1. In general, there are two causes of inefficiency in sequential bargaining: the delay of an agreement and the formation of inefficient subcoalitions. It follows from Theorem 1 (ii) that the delay of an agreement never happens in an SSPE of the random-proposer model, $\Gamma(N, p, \delta)$. Thus, the inefficiency of a payoff allocation arises solely by the formation of a subcoalition. The delay of an agreement may occur in the rejector-proposes model owing to responders' high acceptance thresholds (Chatterjee *et al.*, 1993).

An SSPE of $\Gamma(N, p, \delta)$ is called the *grand coalition SSPE* if the grand coalition N forms, independent of a proposer. It can be shown from Theorem 1 (iii) that the grand coalition SSPE exists in $\Gamma(N, p, \delta)$ if and only if

$$v(N) - \sum_{j \in N} \delta_j v_j \geq v(S) - \sum_{j \in S} \delta_j v_j, \quad \text{for all} \quad S \subset N, \tag{3}$$

$$(1-\delta_i)v_i + p_i \sum_{j \in N} \delta_j v_j = p_i v(N), \quad \forall i \in N, \tag{4}$$

where v_i is the expected payoff of every player i. Equation (4) solves

$$v_i = \frac{\dfrac{p_i}{1-\delta_i}}{\sum_{j \in N} \dfrac{p_j}{1-\delta_j}} v(N),$$

for every $i \in N$. Noting that $\sum_{i \in N} v_i = v(N)$, Equation (3) can be rewritten as

$$\sum_{i \in S} v_i + \sum_{j \in N-S} v_j(1-\delta_j) \geq v(S) \quad \text{for all} \quad S \subset N.$$

Thus, we can prove the following properties of the grand coalition SSPE when all players have the same and sufficiently large discount factors for future payoffs.

Theorem 2 *(Okada, 2011) Suppose that all players have common discount factors δ in the random-proposer model, $\Gamma(N, p, \delta)$. Let p_i be the recognition probability for player i. Then, the following properties hold.*

(i) Every player i's expected payoff of the grand coalition SSPE is equal to $p_i v(N)$.
(ii) The grand coalition SSPE exists for any δ close to 1 if and only if the payoff vector $(p_1 v(N), \cdots, p_n v(N))$ is in the core of (N, v).
(iii) Every player's proposal converges to $(p_1 v(N), \cdots, p_n v(N))$ as δ tends to 1.

When all players are sufficiently patient, the theorem shows that the proportional allocation, $(p_1 v(N), \cdots, p_n v(N))$, of the total value, $v(N)$, according to the recognition probability $p = (p_1, \cdots, p_n)$ is agreed in the grand coalition SSPE, independent of who becomes a proposer. Intuitively, because players can form coalitions freely, the agreement of the grand coalition SSPE should be immune to any coalitional deviation. That is, the grand coalition SSPE payoff should satisfy the core stability. The theorem shows that the grand coalition SSPE exists for any δ close to one if and only if the proportional allocation belongs to the core of the underlying game (N, v). When the recognition probability p is given by the uniform distribution, $(1/n, \cdots, 1/n)$, this condition is equivalent to the condition that the grand coalition N has the highest coalitional value per member:

$$\frac{v(N)}{|N|} \geq \frac{v(S)}{|S|} \quad \text{for all} \quad S \subset N. \tag{5}$$

Chatterjee *et al*. (1993) show that Equation (5) also holds in the rejector-proposes model if and only if the grand coalition N forms, independent of the initial proposer, when players are sufficiently patient. Thus, the efficiency result in non-cooperative coalitional bargaining games is robust with respect to changes in the rules governing the selection of proposers.

The proportional allocation, $(p_1 v(N), \cdot\cdot\cdot, p_n v(N))$, in the grand coalition SSPE is regarded as the (asymmetric) bargaining solution of Nash (1950) for (N, v) that maximizes the product, $\Pi_{i \in N} x_i^{p_i}$, of payoffs over the set of individually rational allocations, where the disagreement payoffs are given by the zero point, $0 = (v(\{i\}))_{i \in N}$. An SSPE is referred to as *asymptotically efficient* if the expected equilibrium payoffs of players converge to an efficient allocation as players become sufficiently patient. Compte and Jehiel (2010) extend Theorem 2 to the case of an asymptotically efficient SSPE in the bargaining game in which only one profitable coalition is allowed to form (like the wage bargaining model in Subsection 2.3). This bargaining game is referred to as a game with the *one-stage property*. When the grand coalition is uniquely efficient, Compte and Jehiel characterize the limit payoff of an asymptotically efficient SSPE as the core-constrained Nash bargaining solution (which they call the *coalitional Nash bargaining solution*) that maximizes the Nash product over the core of the game. The characterization of an SSPE in an n-person game with an empty core is an open problem. Okada (2014) classifies all types of an SSPE in a three-person game in terms of the efficiency level.

Finally, I review the uniqueness results of an SSPE in the random-proposer model. Baron and Ferejohn (1989) establish the uniqueness of an SSPE payoff in a simple-majority voting game when voters are identical in recognition probability and discount factors for future payoffs. Eraslan (2002) extends the result to a q-majority voting game in a general case of unequal recognition probability and time preferences. Eraslan and McLennan (2013) further extend the result to voting games with a general class of winning coalitions. Montero (2006) shows that the nucleolus of a proper weighted majority game is equal to a unique SSPE payoff of the random-proposer model in which the recognition probability is given by the nucleolus itself. Compared with the literature of voting games, the uniqueness of an SSPE payoff has not been well explored for a game in coalitional form. Yan (2002) proves that when the random-proposer model has the one-stage property, every core allocation of a game can be sustained as a unique SSPE payoff if it is used as the recognition probability (after normalization). Okada (2011) shows the generic uniqueness of the asymptotic SSPE payoff for a wage bargaining model. Montero and Okada (2007) show a case of multiple SSPE payoffs in a three-person game with discrete payoffs. The uniqueness problem of an SSPE payoff remains unsolved for a general n-person game in coalitional form.

2.3 Involuntary unemployment: An example

I present an application of the random-proposer model to wage bargaining in a labour market. Since the work of Keynes (1936), there have been theoretical attempts to reconcile involuntary unemployment with the classical Walrasian equilibrium predicting full employment.[9] In a simple example of a labour market, I show that a non-cooperative equilibrium of the coalitional bargaining model can describe both full employment and involuntary unemployment, depending on the model parameters. For a general treatment of the model, see Okada (2011).

[9] Keynes (1936, p. 15) defines involuntary unemployment as follows: “Men are involuntarily unemployed if, in the event of a small rise in the price of wage-goods relatively to the money-wage, both the aggregate supply of labour willing to work for the current money-wage and the aggregate demand for it at that wage would be greater than the existing volume of employment”.

There is one employer, indexed by 1, and there are two identical workers, indexed by 2 and 3. The employer cannot produce any value without workers, and workers cannot work without the employer. For $s = 1, 2, 3$, let $v(s)$ be the total value that the employer can produce when he or she hires $s - 1$ workers. I assume $0 = v(1) \leq v(2) < v(3)$. The value function, v, is monotonically increasing in the number of hired workers and, thus, the full employment outcome where the employer hires all two workers is uniquely Pareto efficient. A situation in which one worker is unemployed is inefficient. The traditional solutions, such as the Walrasian equilibrium and the core in cooperative game theory, predict full employment. The Walrasian equilibrium wage is equal to the worker's reservation wage of zero. The allocation $(v(3), 0, 0)$, where the employer exploits the total surplus, belongs to the core of the underlying cooperative game.

In wage bargaining, the employer and two workers negotiate as to who is employed and how much is paid in terms of wages. Negotiations take place according to the random-proposer rule, with equal recognition probability. Note that all players, including workers, have bargaining power to the extent that they may make proposals with positive probability.

At the start of every round, the employer and two workers have an equal chance of being selected as a proposer. Players have the common discount factor δ for future payoffs, where $0 \leq \delta < 1$. For every $i = 1, 2, 3$, let v_i be the expected payoff of player i in an SSPE. Let S and T be any two coalitions, including i. If player i proposes S with a positive probability in an SSPE, then by the optimality condition (1), it must hold that

$$v(s) - \sum_{j \in S} \delta v_j \geq v(t) - \sum_{j \in T} \delta v_j, \tag{6}$$

where s and t are the number of members in S and T, respectively.

The grand coalition SSPE is called the *full employment* SSPE, in which every player proposes the three-person coalition with probability one. By the payoff Equation (2), it holds that, for all $i = 1, 2, 3$,

$$v_1 = \frac{1}{3}(v(3) - \delta v_2 - \delta v_3) + \frac{2}{3}\delta v_1$$
$$v_2 = \frac{1}{3}(v(3) - \delta v_1 - \delta v_3) + \frac{2}{3}\delta v_2$$
$$v_3 = \frac{1}{3}(v(3) - \delta v_1 - \delta v_2) + \frac{2}{3}\delta v_3.$$

The first equation means that the employer becomes a proposer with probability 1/3 and receives the residual surplus $v(3) - \delta v_2 - \delta v_3$ after he or she pays δv_i to workers $i = 2, 3$. With probability 2/3, he or she becomes a responder and receives payoff δv_1. The other two equations are interpreted in the same way. These equations solve $v_1 = v_2 = v_3 = v(3)/3$. Equation (6) is given by $v(3) - 2v(3)\delta/3 \geq v(2) - v(3)\delta/3$. Thus, the full-employment SSPE exists if and only if

$$\frac{3-\delta}{3} v(3) \geq v(2) \tag{7}$$

and every player receives the same expected payoff, $v(3)/3$. When players are sufficiently patient, they agree to the equal allocation $(v(3)/3, v(3)/3, v(3)/3)$, independent of who

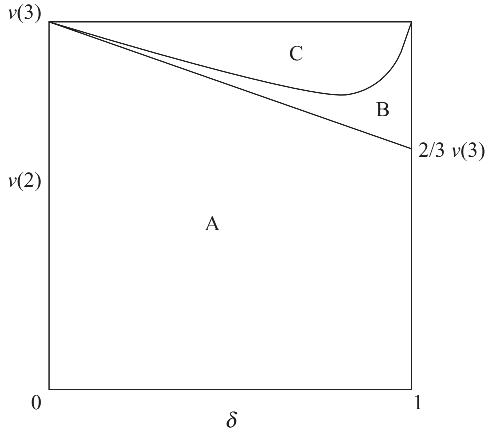

FIGURE 1. The efficiency in a labour market

becomes a proposer. Region A in Figure 1 illustrates the set of parameters, (δ, $v(2)$), for which the full-employment SSPE exists.

An SSPE is called a *partial-employment* SSPE if the probability of full employment is less than one. There are two types of such an equilibrium, depending on whether the probability of full employment is positive or zero.

Suppose that the probability of full employment is positive, but less than one. Let q be the probability that a two-person coalition of the employer and a worker forms. By assumption, $0 < q < 1$. Without any loss of generality, it can be assumed that the employer proposes all feasible coalitions with positive probability.[10] Then, the optimality condition (Equation 1) implies the following equalities

$$v(2)-\delta v_2 = v(2)-\delta v_3 = v(3)-\delta v_2-\delta v_3. \tag{8}$$

Payoff Equation (2) implies

$$v_1 = \frac{1}{3}(v(2)-\delta v_2)+\frac{2}{3}\delta v_1. \tag{9}$$

It follows from Equations (8) and (9) that $v_1 = \dfrac{2v(2)-v(3)}{3-2\delta}$ and $v_2 = v_3 = \dfrac{v(3)-v(2)}{\delta}$. The sum of the three players' expected payoffs is given by

$$v_1 + 2v_2 = (1-q)v(3)+qv(2).$$

It can be seen that this solves

[10] It is proved that all identical workers receive the same expected payoff in every SSPE (see Okada, 2011). Suppose that the three-person coalition and a two-person coalition may be proposed by different players. Even in such a case, it follows from the optimality condition (1) that $v(3) - \delta v_1 - 2\delta v_2 \geq v(2) - \delta v_1 - \delta v_2$, because the three-person coalition may be proposed with a positive probability. By the same reasoning, the opposite inequality must hold, because a two-person coalition may be proposed with a positive probability. Thus, Equation (8) holds in other cases too.

$$q = \frac{2(1-\delta)}{(3-2\delta)\delta} \frac{3v(2)-(3-\delta)v(3)}{v(3)-v(2)}. \tag{10}$$

By Equation (10), it can be seen that the condition of $0 < q < 1$ is equivalent to

$$\frac{3-\delta}{3} v(3) < v(2) < \frac{6-5\delta}{6-3\delta-2\delta^2} v(3). \tag{11}$$

Region B in Figure 1 depicts the set of parameters (δ, $v(2)$) for which the probability of full employment is positive, but less than one in an SSPE.

Finally, suppose that the probability of full employment is zero. In other words, every player proposes a two-person coalition with a single worker. In this SSPE, the employer hires only one worker. Without loss of generality, assume that the employer hires workers 2 and 3, with equal probability. By the optimality condition, it must hold that $v_2 = v_3$. Then, it follows from the payoff equation (Equation 2) that

$$v_1 = \frac{1}{3}(v(2)-\delta v_2) + \frac{2}{3}\delta v_1$$
$$v_2 = \frac{1}{3}(v(2)-\delta v_1) + \frac{1}{6}\delta v_2.$$

The above equations solve $v_1 = \frac{2-\delta}{6-5\delta} v(2)$, $v_2 = v_3 = \frac{2-2\delta}{6-5\delta} v(2)$.[11] It is optimal for the employer to propose a two-person coalition if and only if $v(2) - \delta v_2 \geq v(3) - 2\delta v_2$. Substituting the values of v_1 and v_2 into this condition, it holds that

$$v(2) \geq \frac{6-5\delta}{6-3\delta-2\delta^2} v(3). \tag{12}$$

Region C in Figure 1 depicts the set of parameters, (δ, $v(2)$), for which full employment never occurs in an SSPE.

The analysis of an SSPE has the following implications for the efficiency of a labour market. Full employment is not always possible. The intuition behind this result is that the reservation wages of workers are not zero, but are equal to their discounted expected payoffs. The workers' reservation wages are positive in the sequential bargaining theory, in contrast to the Walrasian equilibrium theory. If the total productivity of two workers is not very high compared to that of a single worker (i.e. the marginal contribution of a worker is not high), then it may be optimal for the employer to hire only one worker.

As Figure 1 shows, the efficiency (region *A*) of wage bargaining depends on two parameters, namely δ and $v(2)$. These parameters represent the discount factor for future payoffs and the productivity of partial employment, respectively. When players are completely impatient ($\delta = 0$), the game has the character of ultimatum bargaining and, thus, the outcome is efficient, independent of the productivity of a single worker. The proposer has complete bargaining power and, thus, he exploits the total value. As δ becomes larger, the range of $v(2)$ attaining efficiency in region *A* becomes smaller. Here, involuntary unem-

[11] In the general case that the employer chooses workers with non-uniform probability, we obtain the same solution from the first equation and $v_1 + 2v_2 = v(2)$.

ployment may occur in regions B and C. In particular, involuntary unemployment occurs with probability one in region C. The boundary between regions B and C is given by the nonlinear function of δ in Equation (12).

It is interesting to examine the limiting outcome of wage bargaining as the discount factor δ tends to one. As Figure 1 shows, the range of $v(2)$ in region C shrinks to an arbitrary small interval as δ becomes close to one, and vanishes at the limit. Note that Equation (12) becomes $v(2) \geq v(3)$ at the limit, which is impossible by the assumption of the value function, v. When the discount factor δ tends to one, only regions A and B are possible in equilibrium. The probability of unemployment in region B given by Equation (10) converges to zero as δ tends to one. Thus, when players are sufficiently patient, the equilibrium outcome of wage bargaining converges to the efficient outcome, independent of $v(2)$. Specifically, in region B, the labour market is "asymptotically" efficient in the sense that efficiency can be attained only at the limit. In contrast, the labour market attains efficiency in region A, independent of the discount factor δ.

The intuition behind the asymptotic efficiency of the labour market can be explained as follows. Because the employer always joins a coalition,[12] his or her expected payoff v_1 satisfies

$$v_1 = p_1\left(v(S) - \sum_{j \in S, j \neq 1} \delta v_j\right) + (1 - p_1)\delta v_1,$$

where S is a coalition that the employer may propose with positive probability. This is rewritten as

$$(1-\delta)v_1 = p_1\left(v(S) - \sum_{j \in S} \delta v_j\right).$$

This equation shows that the employer's expected gain relative to his or her acceptance payoff is equal to the product of his recognition probability and the excess of his or her optimal coalition. By the optimality of an equilibrium coalition, it holds that

$$(1-\delta)v_1 \geq p_1\left(v(N) - \sum_{j \in N} \delta v_j\right) \geq 0.$$

The last inequality holds because the game is super-additive. As δ tends to one, it can be seen that the sum of all players' expected equilibrium payoffs converges to the value of the grand coalition N.

The wage bargaining reveals a variety of payoff allocations in the labour market. Specifically, wages to workers in the two regions, A and B, are structurally different in terms of the limit when players are patient. In region A, the SSPE allocation is the equity allocation ($v(3)/3$, $v(3)/3$, $v(3)/3$) of the full-employment value, and it belongs to the core of the underlying cooperative game, because $2v(2)/3 \leq v(3)$ from Equation (2). The wage in region A is based on *egalitarianism*, in which all individuals should be treated equally. In contrast, in region B, the SSPE allocation is ($2v(2) - v(3)$, $v(3) - v(2)$, $v(3) - v(2)$), and

12 A player is called a *central player* at an SSPE if he or she joins a coalition with probability one. See Okada (2014).

the wage is equal to the workers' marginal contributions. In contrast to egalitarianism, the worker's wage in this case is based on a rule (sometimes called the *competition principle*) that people should be treated according to their efforts and contributions. Here, the employer receives the least payoff in the core. Alternatively, it can be seen that the SSPE allocation in region B maximizes the Nash product, $u_1u_2u_3$, of players' payoffs over the core. In region B, note that the equity allocation does not belong to the core.

It is useful to compare the result of the random-proposer model with that of Stole and Zwiebel (1996), who consider an alternative bargaining model in a labour market. They present a non-cooperative model of intra-firm bargaining, where negotiations take place among the employer and all workers inside a firm. In their model, workers sequentially negotiate for their wages in a pairwise manner with an employer. A given pair of an employer and an employed worker play the Rubinstein's alternating-offers game. If a proposal is rejected, then bargaining may break down with a positive probability. In that event, the worker is removed from the firm, and all other workers, including predecessors, renegotiate with the employer sequentially. Stole and Zwiebel show that a unique subgame perfect equilibrium of their model implements the Shapley value of the underlying cooperative game. In the case of two workers, the employer receives the payoff $\{v(1) + v(2) + v(3)\}/3$. Given the number of workers, Stole and Zwiebel's model always predicts an efficient allocation.

The two models of wage bargaining describe different institutional environments in a labour market. In Stole and Zwiebel's model, all workers are "insiders" in the sense that they are already employed before negotiations. In contrast, the random-proposer model presumes no "insider–outsider" relation among workers. All workers are unemployed at the time of negotiations, and an insider–outsider relation appears only after an agreement of employment is reached. Extending their model, Stole and Zwiebel assume that the employer can choose the optimal number of hired workers, given their intra-firm bargaining outcome. They show that, in contrast to the Walrasian equilibrium level, the employer "over-employs" workers. By comparing the two wage bargaining models, the non-cooperative coalitional bargaining theory clearly shows how institutional aspects in the labour market affect employment and wages.

To summarize, involuntary unemployment may occur, depending on the following economic, psychological and institutional factors: workers' productivity (value function), time preference (discount factor for future payoffs) and negotiation rule (random-proposer). Furthermore, the random-proposer model explains how a worker may be unemployed owing only to misfortune in a random event, and not because of a lack in ability or skill.

2.4 Efficiency with renegotiations

The result of the random-proposer model shows that the efficiency principle underlying the Coase Theorem and the classic cooperative game theory is not always true. However, it may be argued that if an agreement of resource allocation is inefficient, rational agents should be able to renegotiate an efficient agreement. In Okada (2000), I examine whether the possibility of renegotiation is effective for attaining an efficient allocation. In this subsection, I briefly review the result of renegotiation in the random-proposer model.

In a model of renegotiation, it is critical to specify a "disagreement point" (or threat-point) of renegotiations, that is, the outcome that prevails if renegotiations fail, as well as a process of renegotiations. For example, suppose that an inefficient allocation of a coalition, S, is reached in some round, and that players attempt to renegotiate the agreement in the next round. Is the current agreement of an allocation still effectively binding when renegotiations fail? While the answer to this question depends on a legal condition governing the bargaining situation, it may hold in some situation that the ongoing agreement remains effective in the case of unsuccessful renegotiations. This disagreement rule is possibly the implicit assumption behind the intuitive arguments that renegotiations could attain an efficient allocation. I modify the random-proposer model so that it accommodates a process of renegotiations with the aforementioned disagreement rule.

I consider again the random-proposer model. To cover a broad class of repeated bargaining situations, the model is modified so that coalition formation occurs in “real time”, where players receive a flow of payoffs generated in the underlying game (N, v) over periods. When an agreement, (S^t, x^t), of coalition and payoffs is made in some round t, players receive their round-payoffs according to the allocation x^t. Here, (S^t, x^t) is called the *round t-agreement*. If $v(S^t) = v(N)$, then the game stops and the agreement (S^t, x^t) will be implemented in all future rounds. Otherwise, renegotiation starts in the next round $t + 1$. The renegotiation rule is as follows.

Renegotiation rule. *If an agreement,* (S^t, x^t)*, with* $v(S^t) < v(N)$*, is reached in round t, then one player is selected from the player set N in round* $t + 1$ *according to the probability distribution p over N, and he or she proposes a new proposal,* (S^{t+1}, x^{t+1})*, with* $S^t \subset S^{t+1}$ *and* $x^{t+1} \in X^{S^{t+1}}$ *. All members in* S^{t+1} *either accept or reject the new proposal sequentially. If all accept it, then* (S^{t+1}, x^{t+1}) *becomes the round* $(t + 1)$*-agreement and is implemented. Otherwise,* (S^t, x^t) *continues to be the round* $(t + 1)$*-agreement. The same process is repeated in future rounds.*

The random-proposer model with renegotiation is denoted by $\Gamma^r(N, p, \delta)$. Formally, $\Gamma^r(N, p, \delta)$ is represented as an infinite-length extensive game with perfect information, as well as the model $\Gamma(N, p, \delta)$ without renegotiation. Every possible play generates a sequence of agreements, $\{(S^t, x^t)\}_{t=0}^{\infty}$, where (S^t, x^t) is the round t-agreement for each t. Initially, set $S^0 = \emptyset$ and $x^0 = 0$. It is assumed that every player i maximizes his or her expected discounted sum of payoffs.

Theorem 3: *(Okada, 2000) In every SSPE of the random-proposer game,* $\Gamma^r(N, p, \delta)$*, with renegotiations for every discount factor* δ (< 1)*, an agreement of an efficient coalition S with* $v(S) = v(N)$ *is reached in most* $n - 1$ *rounds.*

According to the theorem, if players’ discounted factor for future payoffs is strictly smaller than one, the coalition of players may expand, in general, through renegotiations, and an efficient coalition eventually forms. Intuitively, the equilibrium coalition expands in each round, as long as all incumbent members and new participants are better off by forming a new coalition. The efficiency principle holds true through successive renegotiations under the disagreement rule that prevailing agreements remain effective when renegotiations fail.

Theorem 3 has been extended by several researchers. Seidmann and Winter (1998) prove the theorem in the rejector-proposes model with renegotiation (they call it a “reversible actions” model). Gomes (2005) extends it to a partition function game with externalities. The two restricted properties in the models have been relaxed. Gomes and Jehiel (2005) develop a general set-up where coalitions may break up, and identify a necessary and sufficient condition that guarantees the convergence to efficiency. Hyndman and Ray

(2007) consider non-Markov perfect equilibria for coalitional-form games, and establish the efficiency result.

Finally, note that there is a negative effect of renegotiations in coalitional bargaining. In Okada (2000), I show that, when players are sufficiently patient, they may first propose inefficient subcoalitions. The proposer can exploit the total expected payoffs that all other members of a coalition can gain in future rounds. This *first-mover rent* in renegotiations is missing in the model without renegotiation. When players are sufficiently patient, the first-mover rent becomes large enough to motivate players to propose subcoalitions first. As a result, the process of renegotiation creates "vested interests" for coalition members, which distort the equity of an allocation.

2.5 Externalities and incomplete information

In this last subsection, I briefly review two extensions of the non-cooperative bargaining model: externalities and incomplete information.

Ray and Vohra (1999) consider the rejector-proposes model for a game in partition function form in which the value of a coalition depends on a coalition structure of players. They prove the existence of an SSPE in behaviour strategies, and present an algorithm to generate a coalition structure for a no-delay SSPE. Bloch (1996) considers the same bargaining game with fixed payoff allocations and with no discounting. He shows that any core-stable coalition structure can be attained at an SSPE in pure strategies.

While the partition function has been widely employed as the model of a cooperative game with externalities, there has been insufficient research on how the partition function of a game can be constructed from primitives in an economic situation. The same difficulty applies to the standard model of a game in characteristic function form. A game in strategic form is more appropriate in describing a strategic interdependence among players. Games in characteristic function form and in partition function form are regarded as "reduced models" of a game in strategic form.[13]

A cooperative game in strategic form describes an economic situation in which players can communicate and choose their actions jointly. An agreement of actions is assumed to be enforceable. Widespread externalities prevail and utility may not be transferable. The game covers a wide range of multilateral bargaining problems, including a production economy with externalities, cartel formation of oligopolistic firms, public goods provision, environmental pollution and international alliances.

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 $N = \{1, \cdots, n\}$ is the set of players, and each A_i ($i \in N$) is a finite set of player i's actions. Player i's payoff function, u_i, is a real-valued function on the Cartesian product $A = \Pi_{i \in N} A_i$. For a coalition S of N, 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, c_S, of coalition S is an element of $\Delta(A_S)$ (i.e. a probability distribution on A_S). By abusing the notation, $u_i(c)$ denotes the expected payoff of player i for a correlated action, $c \in \Delta(A)$.

In Okada (2010), I extend the random-proposer model to an n-person cooperative game in strategic form. The negotiation rule is the same as that of the basic model described in Subsection 2.2. It is assumed that once an agreement of a coalition is reached, it becomes

[13] Von Neumann and Morgenstern (1944) construct the characteristic function of a game from its strategic form using the theory of zero-sum two-person games.

binding. A proposal consists of a coalition S and a correlated action $c_S \in \Delta(A_S)$. The discount factor δ of future payoffs is re-interpreted as the continuation probability of negotiations. When a proposal is rejected, negotiations may end, with probability $1 - \delta$, in each round. If this happens, then players not bound by any previous agreements choose their individual actions non-cooperatively, and the game ends. Let $\Gamma(G, p, \delta)$ denote the random-proposer model of G, where p is a recognition probability and δ (< 1) is the probability that negotiations continue.

For a game (N, v) in coalitional form, Theorem 2 shows that the proportional allocation of the total value $v(N)$ according to a recognition probability is agreed in the grand coalition SSPE when players are sufficiently patient. The following theorem generalizes the characterization of an efficient agreement to a game in strategic form.

Theorem 4. *(Okada, 2010) Let $\Gamma(G, p, \delta)$ be the random-proposer model for an* n*-person game G in strategic form. As the continuation probability δ tends to one, the allocation of the grand coalition SSPE for $\Gamma(G, p, \delta)$ converges to the asymmetric Nash bargaining solution of G that solves the maximization problem*

$$\max \sum_{i=1}^{n} p_i \cdot \log[u_i(c) - d_i]$$

$$\text{subject to} \quad (1)\ c \in \Delta(A)$$

$$(2)\ u_i(c) \geq d_i \quad \text{for all} \quad i = 1, \cdots, n,$$

where the weight p_i of player i is given by the recognition probability p, and the disagreement point $d = (d_1, \cdots, d_n)$ is given by a Nash equilibrium payoff of G (in mixed strategies).

Theorem 4 shows that the grand coalition SSPE payoffs of $\Gamma(G, p, \delta)$ must be the (asymmetric) Nash bargaining solution of G when players are patient. The Nash bargaining solution is equal to the proportional allocation, $(p_1 v(N), \cdots, p_n v(N))$, when the bargaining model is applied to a coalitional game (N, v), where p_i is the recognition probability of player i. The disagreement point is given by the zero payoffs, $d = (0, \cdots, 0)$, where the individual payoff $v(\{i\})$ of player i is normalized to zero. In the case of a coalitional game, Theorem 2 shows the necessary and sufficient condition for the grand coalition SSPE payoffs when players are patient. The condition is that the proportional allocation $(p_1 v(N), \cdots, p_n v(N))$ is in the core of the game.

To extend Theorem 2 to the coalitional bargaining problem with externalities, we have to answer the following question: what is an appropriate definition of a core for a cooperative game in strategic form? Traditionally, two core concepts, α-core and β-core, have been studied since the work of Aumann (1961). The key element in the definition of a core with externalities is to formulate how the complementary coalition responds to the deviation of a coalition. The traditional core concepts are defined according to the zero-sum two-person game played by two coalitions. The α-core corresponds to the maximin value of a deviating coalition, and the β-core corresponds to the minimax value. These core concepts presume that the complementary coalition would react by damaging a deviating coalition in the worst way possible. This presumption has been often criticized on the grounds that it allows incredible threats by the complementary coalition.

In contrast to the classic cooperative game approach, in Okada (2010), I show that a non-cooperative bargaining approach can derive a reasonable core concept with an externality. My idea is based on the consistency of a solution, as follows. Suppose that all players accept the Nash bargaining solution as the standard of behaviour. If any coalition of players deviates from this agreement, then all other players are faced with a "new" bargaining problem of how to react. If we hold that the same standard of behaviour should be applied to *every* bargaining problem arising in the game, it should be the case that the remaining players react to the coalitional deviation according to the Nash bargaining solution of their reaction problem. In other words, the Nash bargaining solution of a cooperative game in strategic form 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 behaviour by the complementary coalition. I call this new type of the core the *Nash core* for a cooperative game in strategic form. The argument of consistency naturally leads to the requirement that the Nash bargaining solution should belong to the Nash core. For a precise definition of the Nash core, see Okada (2010).

The Nash core is closely linked to an SSPE of the random-proposer game $\Gamma(G, p, \delta)$. After a coalition of players deviates from the agreement of the grand coalition N, the remaining players negotiate their behaviour in a subgame of the whole game $\Gamma(G, p, \delta)$. It follows from Theorem 4 that an SSPE prescribes the Nash bargaining solution behaviour of the complementary coalition. To make this link clear, the efficiency of the grand coalition SSPE is strengthened so that the largest coalition of active players forms in *every* round of $\Gamma(G, p, \delta)$, both on and off the equilibrium path, independent of the proposer. Such an SSPE is called the *totally efficient* SSPE.

Theorem 5. *(Okada, 2010) Let* $\Gamma(G, p, \delta)$ *be the random-proposer model for an* n*-person cooperative game G in strategic form. If a totally efficient SSPE of* $\Gamma(G, p, \delta)$ *exists for any sufficiently large* δ*, then the asymmetric Nash bargaining solution of G belongs to the Nash core of G.*

The converse of Theorem 5 also holds true under a slightly stronger condition that the Nash bargaining solution belongs to the interior of the strict Nash core. See Okada (2010, Theorem 3) for this result. Thus, it is virtually true that the Nash bargaining solution of a cooperative game in strategic form can be supported by the totally efficient SSPE of the bargaining game $\Gamma(G, p, \delta)$ where the continuation probability δ is close to one if and only if the Nash bargaining solution belongs to the Nash core of G.

The other extension of the non-cooperative bargaining theory is to the case of incomplete information. The aim of this research is to provide a non-cooperative foundation for a cooperative game with incomplete information.

An *n-person Bayesian cooperative game* is represented by $G = (N, \{A_S\}_{S \subset N}, \Omega, \pi, \{u_i, \mathcal{F}_i\}_{i \in N})$. Here, $N = \{1, \cdots, n\}$ is the set of players. For each coalition $S \subset N$ of players, A_S is the set of joint actions for S. Then, Ω is the set of (finite) possible states (or types of players), and π is a probability distribution on Ω, the common prior belief of players. For each $i \in N$, u_i is a real-valued function on $A_S \times \Omega$, and denotes the state-dependent von Neumann–Morgenstern utility function of player i. When player i participates in a joint action $a \in A_S$ as a member of coalition S in state ω, he or she receives utility $u_i(a, \omega)$. A *field* $\mathcal{F}_i$ of Ω represents the information that player i possesses about the state of Ω. For an event, E, which is a subset of Ω, $E \in \mathcal{F}_i$ means that player i knows whether the prevailing state is in event E or in the complementary event, E^c.

A *contract* x_S for coalition S (simply called *S-contract*) is a function from Ω to A_S. For an S-contract x_S, the conditional expected utility of player $i \in S$ relative to $\mathcal{F}_i$ is an $\mathcal{F}_i$-measurable function, $E(u_i(x_S) \mid \mathcal{F}_i) : \Omega \to R$, which is defined by

$$E(u_i(x_S) \mid \mathcal{F}_i)(\omega) = \sum_{\omega' \in I} \pi_{|I}(\omega') u_i(x_S(\omega'), \omega'),$$

for every $\omega \in \Omega$, where $I = I_i(\omega)$ is the *information set*[14] of $\mathcal{F}_i$ containing ω, and $\pi_{|I}(\omega')$ is the posterior belief, given I.

In the Bayesian cooperative game G, players negotiate for coalition formation and contracts. To develop a non-cooperative bargaining theory for the Bayesian cooperative game, the following distinctions are important. The cooperative game G is called *ex ante* if players negotiate before they receive private information about a true state, *interim* if they negotiate after they receive their private information, but not others' information, and *ex post* if they negotiate after the uncertain state becomes publicly known. The game has *verifiable states* if a true state becomes commonly known and is verifiable when a contract is implemented. Otherwise, the game has *unverifiable states*. In this case, a contract should be *incentive compatible* so that players have an incentive to report their private information truthfully. Following the classic works of Harsanyi and Selten (1972) and Wilson (1978), I review recent works on non-cooperative bargaining in the case of interim Bayesian cooperative games with verifiable states. For recent developments in other cases, see Forges *et al*. (2002) and Forges and Serrano (2013).

As in the case of complete information, the core is a fundamental solution concept for a cooperative game with incomplete information. Roughly, an N-contract x^* is in a core of the Bayesian cooperative game G if no coalition S objects to it. In other words, there is no S-contract y_S in which no members of S are better off in y_S than in x^*. However, this definition is incomplete unless one specifies on which event the members of S evaluate an alternative contract y_S. In other words, we need to specify what kind of private information is pooled among the members when they make an objection to the status quo contract x^*. With regard to this issue of information pooling within the coalition, Wilson (1978) considers two extreme situations: no information pooling and full information pooling. He defines a *coarse core* based on the assumption that a coalition may object to an allocation if and only if it is commonly known by its members that they are better off by objecting. In the coarse core, any private information is not shared among the coalition members to organize an objection. As the other polar case, Wilson defines the *fine core* based on the assumption that a coalition may utilize unlimited communication among agents to make an objection. The stability of the fine core is stringent, allowing unlimited communication. Thus, the fine core is a subset of the coarse core, and it may be empty in a standard model of an exchange economy. Vohra (1999) extends Wilson's coarse core to the case of unverifiable types.

Since the work of Wilson (1978), many authors have explored an appropriate definition of the core under incomplete information. Private information may leak through negotiations, and revealed information on uncertain states may change the prospect of agreements. Serrano and Vohra (2007) argue that the non-cooperative equilibrium theory is ideally suited to deal with the question of revealing endogenous information in negotiations. In the case of unverifiable states, they consider a coalitional voting game in which a non-strategic

[14] An information set of $\mathcal{F}_i$ is an element of the finest partition of Ω contained in $\mathcal{F}_i$.

arbitrator without private information chooses an alternative contract, and give non-cooperative support for the *credible core*, taking into account information credibly inferred from the act of objection. Because a coalition can coordinate member voting on any admissible event with the help of the mediator's proposal, the credible core actually coincides with Wilson's fine core in the case of verifiable states.

Elaborating on the work of Serrano and Vohra (2007), in Okada (2012), I consider the rejector-proposes model of the Bayesian cooperative game G with no discounting. A sequential equilibrium of the bargaining game naturally leads to a new type of objection, whereby all members of a coalition are better off after a self-selection event in which a proposal credibly transmits a proposer's private information to responders. The objection based on endogenous information transmission and the corresponding core concept are defined as follows.

Definition 2: *A coalition S has a* signaling objection *to an* N-*contract x if there exist an* S-*contract* y^S, *a member* $i \in S$, *and an event* $E \in \mathcal{F}_i$ *such that*

(i) $E(u_i(y^S) \mid \mathcal{F}_i)(\omega) > E(u_i(x) \mid \mathcal{F}_i)(\omega)$ for all $\omega \in E$,
(ii) $E(u_i(y^S) \mid \mathcal{F}_i)(\omega) \leq E(u_i(x) \mid \mathcal{F}_i)(\omega)$ for all $\omega \notin E$, and
(iii) $E(u_j(y^S)|I_j(\omega) \cap E) > E(u_j(x)|I_j(\omega) \cap E)$ for all $j \in S, j \neq i$, and all $\omega \in E$.

The *signaling core* is the set of all N-contracts to which no coalition has a signaling objection.

Conditions (i) and (ii) enable proposer i's proposal y_S to credibly reveal his or her private information $E \in \mathcal{F}_i$ to responders, because proposer i prefers y^S to the status quo contract x for event E, and not for the complementary event E^c. If responders know this fact, then they can credibly infer that a true state must be in E. Condition (iii) means that all responders accept y_S, given their updated beliefs. It can be shown that the signaling core is a superset of the fine core and a subset of the coarse core. It is known that the fine core, and, thus, the signaling core, exists in an exchange market in which players have quasi-linear utility functions. See Dutta and Vohra (2005) and Okada (2012). The general existence of the signaling core remains an open question.

To obtain a non-cooperative foundation of the signaling core for the Bayesian cooperative game G using the rejector-proposes model, two well-known difficulties need to be overcome. One is the sensitivity of an SSPE outcome to the selection of an initial proposer. An SSPE outcome may not belong to the core, even in the case of complete information, owing to this sensitivity. In the literature of non-cooperative implementation of the core with complete information, several approaches have been proposed to avoid the sensitivity problem. In Okada (1992) and Okada and Winter (2002), we employ the restart rule, according to which there exists an upper bound of successive proposals within each round. If an agreement is not reached within the bound, then the game restarts with the initial proposer. The second difficulty is the multiplicity problem of a sequential equilibrium in non-cooperative sequential bargaining games with incomplete information, owing to an unreasonable belief off the path of equilibrium play. The idea of information revealing underlying the signaling core leads to a refinement of a sequential equilibrium satisfying the property of *self-selection*. Roughly, self-selection means that, given a proposal, every responder updates his or her prior belief off the equilibrium play, and infers that a true state must be in the event that the proposer prefers to object to the status quo allocation. See Okada (2012) for the definition of self-selection refinement.

Theorem 6. *(Okada, 2012) Let G be an* n*-person Bayesian cooperative game, and let* Γ *be the rejector-proposes bargaining game with restart and no discounting. If an* N*-contract x is agreed (with probability one) in a stationary sequential equilibrium of* Γ *that satisfies self-selection, then x belongs to the signaling core of G.*

I also analyse the random-proposer model for a two-person Bayesian cooperative game (Okada, 2013), and extend the characterization and the convergence results of the Nash bargaining solution (Theorem 4) to the case of incomplete information. Specifically, it is shown that the equilibrium proposal of every player converges to the ex-post Nash bargaining solution as the discount factor tends to one if a stationary sequential equilibrium of the bargaining game satisfies the self-selection property and a property called independence of irrelevant types. According to the latter property, the response of every type of a player is independent of proposals to his or her other types.

3. Theory of institution

3.1 The model of institution formation

In this section, I consider institutional foundations for cooperation. Among many institutional factors, the enforceability of agreements is one of the most critical conditions for cooperation. In almost all the bargaining models I reviewed in the previous section, it is assumed that any agreement of cooperation can be enforced once it is reached. How can an agreement of cooperation be enforced? Specifically, how does an institution to enforce an agreement emerge in a society?

To answer these questions, I employ the contractarian point of view that rational individuals voluntarily agree to create an institution for their collective benefit. I examine the possibility of institution formation in a social dilemma situation in which the pursuit of individual interests conflicts with the maximization of social welfare. Public goods provision and common-pool resource management are classic examples of this social dilemma.

In a social dilemma situation, every individual has an incentive to free ride on cooperative actions of others. A solution to the free rider problem seems to create an institution that enforces the maximizing behaviour of group welfare, punishing violators from it. However, since the work of Parsons (1937), there have been pervasive arguments against the contractarian approach to a solution to the free riding problem. It is often argued that rational and selfish individuals have an incentive to free ride on a mechanism that is designed to solve the (first-order) problem of free riding. Despite this negative view of institution formation, Ostrom (1990) investigates empirically the self-governance of common-pool resources, and argues that an effective sanctioning system is a critical factor in the success of governing the commons. Yamagishi (1986) investigates experimentally the voluntary provision of a sanctioning system in public goods games.

Consider the following n-person public goods game. Let $N = \{1, \cdots, n\}$ be the set of players. Every player $i \in N$ has a private endowment, $w > 0$, from which he or she can contribute to a public good. Let $g_i \leq w$ be a contribution of player i. Given a contribution profile $(g_1, \cdots, g_n)$ of players, player i's material payoff is given by

$$u_i(g_1, \cdots, g_n) = w - g_i + a\sum_{i=1}^{n} g_i, \tag{13}$$

where $1/n < a < 1$. Parameter a represents the marginal per capita return (MPCR) from contributing to a public good. Every player is assumed to maximize his or her material payoff. The assumption $1/n < a < 1$ implies that: (i) player i maximizes his or her payoff by contributing nothing ($g_i = 0$), regardless of others' contributions, and, thus, the zero-contribution profile $(0, \cdots, 0)$ is a unique Nash equilibrium; and (ii) the Nash equilibrium is not Pareto efficient, because all players are better off by jointly contributing $g_i = w$.

To solve the problem of public goods provision, some suitable institutional arrangements are needed. As an institutional arrangement, we consider a sanctioning institution that enforces contributions from participants. A process of institution formation is formulated as the following three-stage, non-cooperative game.

Institution formation game:

(i) (Participation stage) Every player announces independently whether to participate in an institution. The institution sanctions members if they do not contribute fully to the public good.

(ii) (Implementation stage) All participants either accept or reject simultaneously and independently the implementation of an institution. The institution is implemented if and only if all participants accept it. The institution is costly. If the institution is implemented, then its costs are shared equally among members. The institutional cost is denoted by $c > 0$.

(iii) (Contribution stage) All players choose their contributions. If an institution is implemented, all members are bound to contribute fully. Other players are free to choose their contributions. If an institution is not implemented, all players are free to choose their contributions.

In the game, every player chooses his or her action, knowing perfectly all players' actions in previous stages. The payoff u_i of player $i \in N$ is given as follows. If an institution with a set $S \subset N$ of participants is implemented, then the payoff is given by

$$u_i = \begin{cases} w - g_i + a\sum_{i=1}^{n} g_i - \dfrac{c}{s} & \text{if } i \in S, \\ w - g_i + a\sum_{i=1}^{n} g_i & \text{if } i \notin S, \end{cases} \tag{14}$$

where $g_i = w$, for all $i \in S$, and s is the cardinality of S. If no institution is implemented, then the payoff of every player $i \in N$ is given by Equation (13).

Interpretation. Institution formation is a complex process, and involves social, political and economic variables. The process is often less structured in the pre-negotiation stages. Informal communication and negotiations play critical roles in the process. Inevitably, any analytically tractable model has to be simple and abstract. The model captures some basic elements in an institution formation process in the real world. First, the sanctioning institution should be voluntary. Any individual should be free from any constraint on his or her own liberty, unless he or she himself is willing to accept it. The model starts with the participation stage in which all individuals voluntarily decide whether to participate in an institution.[15] A non-participant is free from any punishment by the institution. An institution is implemented by the unanimity rule within the set of participants. Every participant has the veto in the implementation of the institution. The voluntary participation in the institution is salient in international negotiations. Second, the process is dynamic. Individuals make their decisions sequentially, updating their expectations of others' behaviour and the prosperity of an institution. The model captures this dynamic process of institution formation in the multi-stage game in which players first announce their (un)willingness to participate in an institution and, thereafter, participants decide to form the institution, knowing the number of participants. If only a few players announce their willingness to participate in an institution, then the institution is likely to fail. Finally, for simplicity, I do not explicitly model the sanctions of an institution, but assume that all members are forced to contribute fully if the institution is successfully implemented.[16] It is implicitly assumed that the members are punished by the institution if they do not contribute fully to the public good, making it optimal for them to contribute fully. Okada (2008) applies the institution formation game to a public goods economy with capital accumulation.

A subgame perfect equilibrium of the institution formation game is analysed using the standard method of backward induction. The equilibrium is classified into two types. A subgame perfect equilibrium of the game is called an *institutional equilibrium* if an institution is implemented on the equilibrium play. Otherwise, it is called a *status quo equilibrium*.

Consider first the contribution stage game. It is clear that non-participants contribute nothing, regardless of the others' contributions, because they are free from punishment. In contrast, all participants contribute fully if an institution is implemented. Otherwise, they contribute nothing in the same manner as non-participants.

Given the equilibrium outcome of the contribution stage, we can now solve the implementation stage. Let $s \leq n$ be the number of participants in an institution. It follows from Equation (14) that every participant receives payoff $asw - \frac{c}{s}$ if the institution is implemented. Otherwise, he or she receives payoff w. Thus, all participants are better off when an institution is implemented than the zero contribution outcome if

$$asw - \frac{c}{s} > w. \tag{15}$$

The smallest integer s^* satisfying Equation (15) is a key factor of the institution formation game, and is called the *minimum institutional size*. If the number s of participants is greater than s^*, then the implementation stage has multiple Nash equilibria under the unanimity rule. In one equilibrium, all participants accept the implementation of an institution. In other equilibria, the institution is not implemented. For example, the action profile in which all participants reject the institution is a Nash equilibrium.

Owing to the multiplicity of the Nash equilibrium in the implementation stage, the institution formation game has many subgame perfect equilibria.

---

15 The voluntary participation problem has been widely studied. See Selten (1973), Palfrey and Rosenthal (1984), Okada (1993), Saijo and Yamato (1999) and Dixit and Olson (2000), among others.

16 The model in this study is a simplified version of the model analysed by Kosfeld *et al*. (2009), where the punishment rule is exogenously introduced. In Okada (1993), the punishment level is determined through collective bargaining within an institution.

Proposition 1: *(Kosfeld* et al.*, 2009) In the institution formation game, there exists an institutional equilibrium in which an institution with s members is implemented if and only if* $s \geq s^*$. *For any number* $s(1 \leq s \leq n)$ *of participants, there exists a status quo equilibrium.*

The proposition has the following implications for the problem of endogenous institution formation. First, the negative view of the social dilemma that cooperation supported by a sanctioning institution fails owing to the free rider problem of the institution is not justified on theoretical grounds. For every integer s greater than or equal to the minimum institutional size s^*, an institution with s participants is implemented in a subgame perfect equilibrium of the game. The institutional equilibrium in the proposition is composed of the strategy profile in which exactly s players participate in and implement the institution, while all institutions of different sizes are rejected.

Second, institution formation is not always possible. For every number of participants, the status quo equilibrium always exists. In equilibrium, an institution is rejected in the implementation stage, regardless of the number of possible participants.

Third, the success of institution formation depends on players' expectations about other players' behaviour. To implement an institution, players have to solve two kinds of coordination problems. The first problem is with respect to the size of the institution. The second problem is with respect to who become members, and who will stay out.

The multiplicity of a subgame perfect equilibrium in Proposition 1 is solved in terms of the institutional size if a refinement of a strict equilibrium is applied.[17] A subgame perfect equilibrium of the institution formation game is called strict if it induces a strict Nash equilibrium on every stage game, both on and off the equilibrium play.

Proposition 2: *(Kosfeld* et al.*, 2009) Let* s^* *be the minimum institutional size given by Equation (15). The institution formation game has a unique strict subgame perfect equilibrium in terms of the institutional size. In this equilibrium, exactly* s^* *players participate in an institution and the institution is implemented.*

The intuition behind the proposition is as follows. If the number of participants is larger than the minimum institutional size s^*, then the requirement of strictness selects a unique Nash equilibrium in the implementation stage, where all participants accept an institution. Owing to the equilibrium selection in the implementation stage, an institution with more than s^* participants is not implemented in equilibrium because every member is better off if he or she opts out. If the number of participants is equal to s^*, then no participant has an incentive to opt out because, if he or she does so, the institution is not implemented. Thus, only an institution with s^* participants is possible in equilibrium. Moreover, the status quo equilibrium is not a strict subgame perfect equilibrium because every participant is indifferent to the participation decision. If $s^* \neq n$, then players are divided into two proper subsets: those who voluntarily implement an institution, and, thus, contribute to the public good, and those who do not participate and do not contribute. The equilibrium prediction of strictness is unfavorable for symmetry, equality and efficiency.

Note that the minimum institutional size, s^*, in Equation (15) is defined under the standard assumption that players maximize their material payoffs. If they have social preferences such that institutional members dislike payoff inequality against free riders, then the minimum institutional size may become larger. As a result, an institution with free

[17] A Nash equilibrium of a strategic form game is called strict if every player has a unique best response to the other players' strategies.

riders may be rejected. Kosfeld *et al*. (2009) present an equilibrium analysis of the institution formation game under social preferences.

To conclude this subsection, I briefly review the experimental results for endogenous institution formation. Kosfeld *et al*. (2009) report experimental findings for the institution formation game. The experiments were conducted at the CREED laboratory at the University of Amsterdam. Subjects played 20 rounds of the institution formation game with the same group of four players (partner matching). Two experimental treatments were implemented with different minimum institutional sizes, $s^* = 2, 3$. The MPCR of public goods, $a = 0.4$ and 0.65, were chosen to provide $s^* = 2$ and 3, respectively. The main experimental findings are as follows. First, subjects successfully establish institutions. In both treatments, they implement institutions between 43 and 61% of the cases. Second, the majority (on average, approximately 70%) of the implemented institutions are the largest ones, with four members. Third, the possibility of institution formation has a positive effect on contributions to the public good.

Sutter *et al*. (2010) investigate whether subjects choose a voluntary contribution mechanism with rewards or punishments. They find that group members typically vote for the reward option, and that the endogenous institutional choice has a significant and positive effect on the level of cooperation compared with the same exogenously implemented institution. In the institution formation game, we do not explicitly formulate a sanction mechanism. The model can be applied to any kind of mechanism that enforces a collective action on group members. Traulsen *et al*. (2012) investigate an institutional choice of subjects between two punishment mechanisms: a decentralized mechanism (called a peer punishment) and a centralized mechanism (called a pool punishment). They find that the majority of subjects choose the centralized punishment in the presence of second-order punishment (the punishment of cooperators who do not punish).

3.2 Decentralized institution of cooperation

While the model of institution formation in the previous subsection is formulated to describe a situation in which the institution has a centralized enforcement agency, such as police or courts, the model can be applied, in principle, to any kind of an institution with an enforcement mechanism. An institution may be decentralized in the sense that the enforcement of an agreement is implemented by players themselves through mutual monitoring and punishment. In the literature, repeated game strategy, such as the trigger strategy, has been studied intensively as a particular form of such a decentralized institution. A subgame perfect equilibrium in a repeated game is self-binding, because no player has an incentive to deviate from it. In this subsection, I regard a subgame perfect equilibrium of a repeated game as a decentralized institution, and consider how such an institution of cooperation is voluntarily established among players.

The voluntary formation of an institution is particularly relevant in the theory of repeated games. The folk theorem of the repeated game states that if individuals are patient, every individually rational outcome in a stage game can be sustained as a subgame perfect equilibrium of the repeated game. A well-known drawback of the folk theorem is that the set of equilibrium outcomes is plethoric. In the public good game in the previous subsection, any group larger than or equal to the minimum institutional size s^* given by Equation (15) can be sustained in a subgame perfect equilibrium of the repeated game if players are patient. The zero contribution outcome is also sustained in a subgame perfect equilibrium. However, which subgame perfect equilibrium is played remains an open problem. The model of institution formation provides an answer to the equilibrium selection problem in the repeated game.

Consider again the public goods game in Equation (13). Without loss of generality, assume that each player $i \in N$ has a binary choice: a zero contribution ($g_i = 0$) or full contribution ($g_i = \omega$). The zero contribution is referred to as defection (D), while full contribution is referred to as cooperation (C). The institutional cost, c, is assumed to be zero. The public goods game is repeated for infinitely many rounds under the condition of perfect information that every player knows the choices of all players in all past rounds. Players have common discount factors, δ ($0 \leq \delta < 1$), for their future payoffs.

A group of players is called *individually rational* if its size is larger than or equal to the minimum institutional size s^*. If all members cooperate in an individually rational group, then they are better off than in the defection equilibrium, where all players defect. For an individually rational group, define the *group-trigger strategy* such that all group members cooperate if and only if they cooperated in all past rounds, and all non-members always defect. Note that non-members are not punished when they deviate. If players are sufficiently patient, the group-trigger strategy constitutes a subgame perfect equilibrium of the repeated public goods game. As a credible agreement within an individually rational group, I focus on the group-trigger strategy.[18] In what follows, the group-trigger strategy for a group S is simply referred to as the S-trigger strategy.

Maruta and Okada (2012) consider group formation in the repeated provision game of public goods by applying the institution formation game in Subsection 3.1 and the renegotiation model in Subsection 2.4. Players attempt to form (and reform) a group of cooperation in every round. The members of a group are bound to implement the group-trigger strategy. The group-trigger strategy is subject to renegotiation.

A process of group formation is formulated as follows. The game in each round t has a state ω_t, which is either a *negotiation state* or a *non-negotiation state*. If ω_t is a negotiation state, then players have an opportunity to negotiate a group, $S_{t-1} \subset N$ (possibly $S_{t-1} = \emptyset$), that has been formed before round t. The members of S_{t-1} have already agreed to play the group-trigger strategy. In this case, we write $\omega_t = S_{t-1}$. If ω_t is a non-negotiation state, then players do not have an opportunity to form a group, and the game in round t is identical to the original provision game where n players choose their contributions to public goods independently. The non-negotiation state is denoted by ω^*.

When $\omega_t = S_{t-1} \subset N$, the game in round t has the following three stages:

(i) (Participation stage) All players outside S_{t-1} decide independently and simultaneously whether to participate in S_{t-1}. Let P_t be the set of all new participants.

(ii) (Implementation stage) If the expanded group, $S_{t-1} \cup P_t$, is individually rational, then all members of the group either accept or reject the new group sequentially, according to some fixed order. The choice of an order never affects the result. The agreement is made by unanimity. If all accept it, then the new group, $S_t = S_{t-1} \cup P_t$, forms, and its members agree to implement the S_t-trigger strategy, replacing the (ongoing) S_{t-1}-trigger strategy. Otherwise, S_{t-1} remains as S_{t-1}. This rule ($S_t = S_{t-1}$) also applies to the case in which the expanded group, $S_{t-1} \cup P_t$, is not individually rational. In this case, the group-trigger strategy cannot be a self-binding agreement.

(iii) (Action stage) All players in N choose their actions simultaneously.

[18] This causes no loss of generality. The same result holds for any subgame perfect equilibrium strategy that sustains cooperation in every individually rational group.

The transition of states is governed by the following rule, where $\omega_t = S_{t-1}$ in the first three cases:

$$\omega_{t+1} = \begin{cases} S_t & \text{if a new group } S_t \text{ forms and all members of } S_t \text{ cooperate,} \\ S_{t-1} & \text{if no new group forms and all members of } S_{t-1} \text{ cooperate,} \\ \omega^* & \text{if at least one member of the group } (S_t \text{ or } S_{t-1}) \text{ defects,} \\ \omega^* & \text{if } \omega_t = \omega^*. \end{cases} \tag{16}$$

Let Γ denote the repeated public goods game with group formation defined above. Formally, Γ is formulated as a dynamic game with state variables given in extensive form. Every player has perfect information about other players' choices in all past stages.

For a (pure) strategy profile $\sigma = (\sigma_1, \cdots, \sigma_n)$ for players in Γ, let S_t be a group formed in period $t = 1, 2, \cdots$ on the play of σ. The sequence $\{S_t\}_{t=1}^{\infty}$ is called a *group sequence* of σ. By the rule of Γ, a group sequence $\{S_t\}$ is (weakly) monotonically increasing, and there exists some integer, m, such that $S_t = S_{t+1}$, for all $t \geq m$. Such a group, S_m, is called an *absorbing group* of σ. Because $\{S_t\}$ is monotonically increasing, an absorbing group is unique.

As in the repeated game, the multiple equilibrium problem arises in Γ. In particular, the folk theorem of the repeated game also applies to Γ. That is, all individual rational outcomes in the public goods game are sustained in a subgame perfect equilibrium of Γ, if players are patient. Intuitively, Γ generates all possible plays in the standard repeated game of the provision game, because Γ is identical to the repeated provision game if no players participate in a group. For every individually rational group S, the following strategy profile is a subgame perfect equilibrium of Γ attaining cooperation in S. Players never participate in any group, and they behave in the action stage of every period according to the S-trigger strategy when no group forms. To overcome the multiple equilibrium problem, I focus on a Markov-perfect equilibrium of Γ, as in the non-cooperative coalitional bargaining theory reviewed in Section 2. In addition to the Markov property, we need to refine the strict equilibrium to eliminate the "status quo equilibrium" with no participants, as in the institution formation game in Subsection 3.1.

Definition 3: *A pure strategy profile* $\sigma^* = (\sigma_1^*, \cdots, \sigma_n^*)$ *of* Γ *is called a solution of* Γ *if it satisfies the following properties.*

(i) (subgame perfection) σ^* is a subgame perfect equilibrium of Γ.
(ii) (Markov property) The strategy of every player i induced by σ_i^* in every round t depends only on a state variable ω_t.[19]
(iii) (strictness) Let $\{S_t^*\}_{t=1}^{\infty}$ be the group sequence of σ^*. Then, each S_t^* is attained by a strict Nash equilibrium (if any) of the participation stage game induced by σ^* in round t.[20]

[19] I introduce a slightly stronger requirement that the strategy of every player i in round t depends only on the size of group S_{t-1} in ω_t, and on whether $i \in S_{t-1}$.

[20] Note that the strictness is required only for the participation game on the equilibrium path where players' payoffs are defined under the condition that all future plays are given according to σ^*.

We are now in a position to present the following theorem. A group of players is said to be (Pareto) efficient if the outcome that all group members cooperate and all non-members defect is efficient.

Theorem 7. *(Maruta and Okada, 2012) Let* Γ *be the repeated public goods game with group formation where players are sufficiently patient. Then, there exists a solution of* Γ *with an absorbing group,* S^**, if and only if* S^* *is an efficient and individually rational group.*

The theorem shows that an efficient and individually rational group of players necessarily forms in the repeated public good game when players have the opportunity to reform a group in every period. The possibility of renegotiation enables the group formation device to select an efficient group as the absorbing state of a solution. The result generalizes the efficiency principle (Theorem 6) in coalitional bargaining to the context of a repeated game.

The intuition behind Theorem 7 can be explained as follows. Suppose that some group, S (possibly the empty set), of players has already formed in past rounds and the S-trigger strategy is in effect. Then, all members of S cooperate and all non-members free ride. For group S to expand to a new group, $T(\supset S)$, it must hold that all incumbent members of S and all new members in $T - S$ become better off by forming T than by forming S. While every incumbent member of S is better off whenever S is expanded, the new members are better off than those free riding on S only if the size of the new group, T, is larger than a certain threshold. Here, group T is called a *cooperative group given S*. In addition, group S is called a *maximal cooperative group* if there exists no cooperative group given S. Because it is impossible that all players are better off than when a maximal cooperative group forms, group S is efficient and individually rational if and only if it is a maximal cooperative group. A key lemma states that in every solution with a group sequence, $\{S_t\}_{t=1}^{\infty}$, S_{t-1} is expanded to S_t if there exists some cooperative group S given S_{t-1}. The two equilibrium refinements, the Markov property and strictness, are critical to the lemma. If the absorbing group of a solution is not efficient, then it is not a maximal cooperative group. That is, there exists some cooperative group given the absorbing group. Then, the lemma implies that the absorbing group will be expanded, which is a contradiction.

Finally, I discuss two aspects in the decentralized model of institution formation reviewed in this subsection. First, the model has the property that if renegotiations of group formation fail, then the status quo group prevails. This means that the threat point of renegotiations is the current agreement of the group-trigger strategy. If renegotiation is successful, then the status quo group is expanded and both incumbent members and new participants will be bound to the new group-trigger strategy. The group can only expand in the model. A central issue in the efficiency result is how group members are motivated to keep cooperating during the process of renegotiations. This incentive problem is solved by the self-binding agreement of a group-trigger strategy and by the assumption that group members never leave a group once it is formed. Second, players can renegotiate the ongoing group, but cannot renegotiate the punishment of the trigger strategy off the equilibrium play. This approach contrasts with the literature on a renegotiation-proof equilibrium (Farrell and Maskin, 1989), which assumes players can renegotiate at any time, both on and off the equilibrium play. The literature shows that a renegotiation-proof equilibrium is not always efficient. Theorem 7 shows us that renegotiations for ongoing groups, with players' commitment not to renegotiate punishments, are efficient in social dilemma situations.

4. Concluding remarks

I have reviewed recent works on game theoretical analyses on cooperation and institutions in the framework of non-cooperative coalitional bargaining theory. The possibility of cooperation is determined by economic, psychological and institutional factors. Specifically, I have re-examined the efficiency principle, a widely held view in economics, under the strategic behaviour of coalition formation. The analysis shows that the classic solutions of the Nash bargaining solution and the core are closely related to the efficiency of negotiations. When individuals are sufficiently patient, the efficient coalition of all individuals forms in equilibrium if and only if the Nash bargaining solution belongs to the core. If individuals can renegotiate inefficient agreements, then coalitions may expand and eventually reach the grand coalition. Lastly, I extended the non-cooperative coalitional bargaining theory to include the wider problems of externalities, incomplete information, and institution formation.

Seventy years after the foundation provided by von Neumann and Morgenstern in 1944, game theory continues to be an active research field in various disciplines beyond economics. It is hoped that game theoretical investigations will further our understanding of human behaviour and of the sustainability of human society through cooperation.

Acknowledgements

This paper is based on my presidential address delivered to the annual meeting of the Japanese Economic Association, held at Seinan Gakuin University on 12 October 2014. I am grateful to an anonymous referee, Werner Güth, Hartmut Kliemt, Toshiji Miyakawa, Nozomu Muto, Arno Riedl, Yasuhiro Shirata, and participants in the summer schools at the Max Planck Institute of Economics, Jena and the Frankfurt School of Management and Finance in 2012 for their useful comments. Financial support from Hitotsubashi University is gratefully acknowledged.

Final version accepted 16 November 2014.

REFERENCES

Adachi, T. and Y. Watanabe (2007) “Ministerial Weights and Government Formation: Estimation Using a Bargaining Model”, *Journal of Law, Economics, and Organization*, Vol. 24, pp. 95–119.

Albers, W. (1975) “Zwei Lösungskonzepte Für Kooperative Mehrpersonenspiele, Die Auf Anspruchsniveaus Der Spieler Basieren”, *OR-Verfahren*, Vol. XXI, Meisenheim, pp. 1–13.

Aumann, R. (1961) “The Core of a Cooperative Game without Side Payments”, *Transactions of the American Mathematical Society*, Vol. 98, pp. 539–552.

Bandyopadhyay, S. and K. Chatterjee (2006) “Coalition Theory and Its Applications: A Survey”, *The Economic Journal*, Vol. 116, pp. 136–155.

Banks, J. and J. Duggan (2000) “A Bargaining Model of Collective Choice”, *American Political Science Review*, Vol. 94, pp. 73–88.

Baron, D. and J. Ferejohn (1989) “Bargaining in Legislatures”, *American Political Science Review*, Vol. 83, pp. 1181–1206.

Bloch, F. (1996) “Sequential Formation of Coalitions in Games with Externalities and Fixed Payoff Division”, *Games and Economic Behavior*, Vol. 14, pp. 90–123.

Chatterjee, K., B. Dutta, D. Ray and K. Sengupta (1993) “A Noncooperative Theory of Coalitional Bargaining”, *Review of Economic Studies*, Vol. 60, pp. 463–477.

Coase, R. H. (1960) "The Problem of Social Cost", *Journal of Law and Economics*, Vol. 3, pp. 1–44.

Compte, O. and P. Jehiel (2010) "The Coalitional Nash Bargaining Solution", *Econometrica*, Vol. 78, pp. 1593–1623.

Cooter, R. D. (1989) "The Coase Theorem", in J. Eatwell, M. Milgate and P. Newman, eds, *The New Palgrave: Allocation, Information, and Markets*, London: Macmillan, pp. 64–70.

Dixit, A. and M. Olson (2000) "Does Voluntary Participation Undermine the Coase Theorem?" *Journal of Public Economics*, Vol. 76, pp. 309–335.

Dutta, B. and R. Vohra (2005) "Incomplete Information, Credibility, and the Core", *Mathematical Social Sciences*, Vol. 50, pp. 148–165.

Eraslan, H. (2002) "Uniqueness of Stationary Equilibrium Payoffs in the Baron–Ferejohn Model", *Journal of Economic Theory*, Vol. 103, pp. 11–30.

Eraslan, H. and A. McLennan (2013) "Uniqueness of Stationary Equilibrium Payoffs in Coalitional Bargaining", *Journal of Economic Theory*, Vol. 148, pp. 2195–2222.

Evans, R. (1997) "Coalitional Bargaining with Competition to Make Offers", *Games and Economic Behavior*, Vol. 19, pp. 211–220.

Farrell, J. P. and E. Maskin (1989) "Renegotiation in Repeated Games", *Games and Economic Behavior*, Vol. 1, pp. 327–360.

Forges, F., E. Minelli and R. Vohra (2002) "Incentives and the Core of An Exchange Economy: A Survey", *Journal of Mathematical Economics*, Vol. 38, pp. 1–41.

Forges, F. and R. Serrano (2013) "Cooperative Games with Incomplete Information: Some Open Problems", *International Game Theory Review*, Vol. 15, No. 2, pp. 1–17.

Gomes, A. (2005) "Multilateral Contracting with Externalities", *Econometrica*, Vol. 73, pp. 1329–1350.

Gomes, A. and P. Jehiel (2005) "Dynamic Processes of Social and Economic Interaction: On the Persistence of Inefficiencies", *Journal of Political Economy*, Vol. 113, pp. 626–667.

Gul, F. (1989) "Bargaining Foundations of the Shapley Value", *Econometrica*, Vol. 57, pp. 81–95.

Harsanyi, J. C. (1974) "An Equilibrium-Point Interpretation of Stable Sets and a Proposed Alternative Definition", *Management Science*, Vol. 20, pp. 1472–1495.

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

Hart, S. and A. Mas-Colell (1996) "Bargaining and Value", *Econometrica*, Vol. 64, pp. 357–380.

Horniaček, M. (2008) "Negotiation, Preferences over Agreements, and the Core", *International Journal of Game Theory*, Vol. 37, pp. 235–249.

Hurwicz, L. (1960) "Optimality and Informationally Efficiency in Resource Allocation Processes", in K. J. Arrow, S. Karlin and P. Suppes, eds, *Mathematical Methods in the Social Sciences*, Stanford: Stanford University Press, pp. 27–46.

Hyndman, K. and D. Ray (2007) "Coalition Formation with Binding Agreements", *Review of Economic Studies*, Vol. 74, pp. 1125–1147.

Kawamori, T. (2008) "A Repeated Coalitional Bargaining Model", *Mathematical Social Sciences*, Vol. 55, pp. 292–298.

Keynes, J. M. (1936) *The General Theory of Employment, Interest, and Money*, London: Macmillan Press.

Kosfeld, M., A. Okada and A. Riedl (2009) "Institution Formation in Public Goods Games", *American Economic Review*, Vol. 99, pp. 1335–1355.

Laruelle, A. and F. Valenciano (2008) "Noncooperative Foundations of Bargaining Power in Committees and the Shapley–Shubik Index", *Games and Economic Behavior*, Vol. 63, pp. 341–353.

Maruta, T. and A. Okada (2012) "Dynamic Group Formation in the Repeated Prisoner's Dilemma", *Games and Economic Behavior*, Vol. 74, pp. 269–284.

Miyakawa, T. (2009) "Existence and Efficiency of a Stationary Subgame-Perfect Equilibrium in Coalitional Bargaining Models with Nonsuperadditive Payoffs", *Economic Theory*, Vol. 39, pp. 291–306.

Moldovanu, B. and E. Winter (1994) "Core Implementation and Increasing Returns to Scale for Cooperation", *Journal of Mathematical Economics*, Vol. 23, pp. 533–548.

—— and —— (1995) "Order Independent Equilibria", *Games and Economic Behavior*, Vol. 9, pp. 21–34.

Montero, M. (2002) "Non-Cooperative Bargaining in Apex Games and the Kernel", *Games and Economic Behavior*, Vol. 41, pp. 309–321.

—— (2006) "Noncooperative Foundations of the Nucleolus in Majority Games", *Games and Economic Behavior*, Vol. 54, pp. 380–397.

Montero, M. and A. Okada (2007) "Riskless versus Risky Bargaining Procedures: The Aumann–Roth Controversy Revisited", *Japanese Economic Review*, Vol. 58, pp. 294–302.

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

—— (1951) "Non-Cooperative Games", *Annals of Mathematics*, Vol. 54, pp. 286–295.

—— (1953) "Two-Person Cooperative Games", *Econometrica*, Vol. 21, pp. 128–140.

Okada, A. (1992) "Noncooperative Bargaining and the Core of an N-Person Characteristic Function Game", *Control and Cybernetics*, Vol. 21, pp. 231–250.

—— (1993) “The Possibility of Cooperation in an N-Person Prisoners’ Dilemma with Institutional Arrangements”, *Public Choice*, Vol. 77, pp. 629–656.
—— (1996) “A Noncooperative Coalitional Bargaining Game with Random Proposers”, *Games and Economic Behavior*, Vol. 16, pp. 97–108.
—— (2000) “The Efficiency Principle in Non-Cooperative Coalitional Bargaining”, *Japanese Economic Review*, Vol. 51, pp. 34–50.
—— (2008) “The Second-Order Dilemma of Public Goods and Capital Accumulation”, *Public Choice*, Vol. 135, pp. 165–182.
—— (2010) “The Nash Bargaining Solution in General *N*-Person Cooperative Games”, *Journal of Economic Theory*, Vol. 145, pp. 2356–2379.
—— (2011) “Coalitional Bargaining Games with Random Proposers: Theory and Application”, *Games and Economic Behavior*, Vol. 73, pp. 227–235.
—— (2012) “Non-Cooperative Bargaining and the Incomplete Informational Core”, *Journal of Economic Theory*, Vol. 147, pp. 1165–1190.
—— (2013) “A Non-Cooperative Bargaining Theory with Incomplete Information: Verifiable Types”, Discussion Paper 2013–15, Hitotsubashi University.
—— (2014) “The Stationary Equilibrium of Three-Person Coalitional Bargaining Games with Random Proposers: A Classification”, *International Journal of Game Theory*, Vol. 43, pp. 953–973.
Okada, A. and E. Winter (2002) “A Noncooperative Axiomatization of the Core”, *Theory and Decision*, Vol. 53, pp. 1–28.
Oliver, P. (1980) “Rewards and Punishments as Selective Incentives for Collective Action: Theoretical Investigations”, *American Journal of Sociology*, Vol. 85, pp. 1356–1375.
Osborne, M. J. and A. Rubinstein (1990) *Bargaining and Markets*, New York: Academic Press.
Ostrom, E. (1990) *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge: Cambridge University Press.
Palfrey, T. and H. Rosenthal (1984) “Participation and the Provision of Discrete Public Goods: A Strategic Analysis”, *Journal of Public Economics*, Vol. 24, pp. 171–193.
Parsons, T. (1937) *The Structure of Social Action*, New York: MacGraw-Hill.
Pérez-Castrillo, D. and D. Wettstein (2001) “Bidding for the Surplus: A Non-Cooperative Approach to the Shapley Value”, *Journal of Economic Theory*, Vol. 100, pp. 274–294.
Perry, M. and P. J. Reny (1994) “A Noncooperative View of Coalition Formation and the Core”, *Econometrica*, Vol. 62, pp. 795–817.
Ray, D. (2007) *A Game-Theoretic Perspective on Coalition Formation*, Oxford: Oxford University Press.
Ray, D. and R. Vohra (1999) “A Theory of Endogenous Coalition Structures”, *Games and Economic Behavior*, Vol. 26, pp. 286–336.
—— and —— (2014) “Coalition Formation”, in P. Young and S. Zamir, eds, *Handbook of Game Theory*, Vol. 4, Amsterdam: North-Holland, pp. 239–326.
Rubinstein, A. (1982) “Perfect Equilibrium in a Bargaining Model”, *Econometrica*, Vol. 50, pp. 97–109.
Saijo, T. and T. Yamato (1999) “A Voluntary Participation Game with a Non-Excludable Public Good”, *Journal of Economic Theory*, Vol. 84, pp. 227–242.
Seidmann, D. J. and E. Winter (1998) “A Theory of Gradual Coalition Formation”, *Review of Economic Studies*, Vol. 65, pp. 793–815.
Selten, R. (1973) “A Simple Model of Imperfect Competition, Where 4 Are Few and 6 Are Many”, *International Journal of Game Theory*, Vol. 2, pp. 141–201.
—— (1981) “A Noncooperative Model of Characteristic-Function Bargaining”, in V. Boehm and H. Nachthanp, eds, *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*, Mannheim: Bibliographisches Institut, pp. 131–151.
Serrano, R. (1995) “A Market to Implement the Core”, *Journal of Economic Theory*, Vol. 67, pp. 285–294.
Serrano, R. and R. Vohra (1997) “Non-Cooperative Implementation of the Core”, *Social Choice and Welfare*, Vol. 14, pp. 513–525.
—— and —— (2007) “Information Transmission in Coalitional Voting Games”, *Journal of Economic Theory*, Vol. 134, pp. 117–137.
Smith, A. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*, E. Cannan, ed., Chicago: University of Chicago Press.
Snyder, J. M. Jr, M. M. Ting and S. Ansolabehere (2005) “Legislative Bargaining under Weighted Voting”, *American Economic Review*, Vol. 95, pp. 981–1004.
Stole, L. A. and J. Zwiebel (1996) “Intra-Firm Bargaining under Non-Binding Contracts”, *Review of Economic Studies*, Vol. 63, pp. 375–410.
Sutter, M., S. Haigner and M. G. Kocher (2010) “Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations”, *Review of Economic Studies*, Vol. 77, pp. 1540–1566.
Sutton, J. (1986) “Non-Cooperative Bargaining Theory: An Introduction”, *Review of Economic Studies*, Vol. 53, pp. 709–724.

Traulsen, A., T. Röhl and M. Milinski (2012) "An Economic Experiment Reveals That Humans Prefer Pool Punishment to Maintain the Commons", *Proceedings of the Royal Society B*, Vol. 279, pp. 3716–3721.
Vohra, R. (1999) "Incomplete Information, Incentive Compatibility, and the Core", *Journal of Economic Theory*, Vol. 86, pp. 123–147.
von Neumann, J. and O. Morgenstern (1944) *Theory of Games and Economic Behavior*, Princeton: Princeton University Press.
Wilson, R. (1978) "Information, Efficiency, and the Core of an Economy", *Econometrica*, Vol. 46, pp. 807–816.
Yamagishi, T. (1986) "The Provision of a Sanctioning System as a Public Good", *Journal of Personality and Social Psychology*, Vol. 51, pp. 110–116.
Yan, H. (2002) "Noncooperative Selection of the Core", *International Journal of Game Theory*, Vol. 31, pp. 527–540.