The γ-Core and Coalition Formation[1]

Parkash Chander[2]

CORE Discussion Paper 2003/46

(This version: January 2005)

Abstract

This paper reinterprets the γ-core (Chander and Tulkens (1995, 1997)) and justifies it as well as its prediction that the efficient coalition structure is stable in terms of the coalition formation theory. We assume that coalitions can freely merge or break apart and are farsighted (that is, it is the final and not the immediate payoffs that matter to the coalitions). We then show that subsequent to a deviation by a coalition, it is ex post optimal for the nonmembers to break apart into singletons, as is assumed in the definition of the γ- characteristic function, and the grand coalition is a stable coalition structure. Our analysis not only justifies the γ-core in terms of the coalition formation theory but also shows that in order to rule out the stability of the efficient coalition structures one must either place some exogenous restrictions on the process of coalition formation or assume that the coalitions are not farsighted.

JEL Classification numbers: C71, C72, D62.

Keywords: core, characteristic function, strategic games, coalition formation.

---

[1] This work was completed during my visits to CORE in 2002 and 2003. I wish to thank the seminar participants for helpful comments and CORE for its hospitality. The paper was earlier presented at the conference on Game Theory and Its Applications held in Mumbai in 2003. I am thankful to Anindya Bhattacharya and Henry Tulkens for their comments and suggestions.

[2] Indian Statistical Institute, and National University of Singapore, 1 Arts Link, AS 2, Level 6, Singapore 117570. Email: ecsparka@nus.edu.sg.

The γ-Core and Coalition Formation

1. Introduction

The concept of a characteristic function which specifies the worth of each coalition is central to the theory of cooperative games. The worth of a coalition is what it can achieve on its own without the cooperation of the nonmembers. If there are no externalities, i.e., if the payoffs to the members of a coalition do not depend on the actions of the nonmembers, then the worth can be defined without specifying the actions of the nonmembers. But if externalities are present, then in order to calculate the worth of a coalition one must also predict the actions of the nonmembers. This has been however a disputed issue and alternative assumptions in this respect lead to different concepts of characteristic functions such as the α-, β-, and γ - characteristic functions.[3]

The γ- characteristic function was introduced most recently by Chander and Tulkens (1995, 1997) based on an assumption concerning the behavior of the nonmembers which is more plausible than those considered previously. They consider a game in strategic form with transferable utility in which it is efficient for the grand coalition to form and choose the strategy profile that maximizes the joint surplus. They assume that when a coalition deviates it neither takes as given the strategies of its complement, as in the case of strong and coalition proof Nash equilibria, nor presumes that the complement would follow minimax or maximin strategies, as in the case of α - and β - characteristic functions. Instead, it assumes that the nonmembers will not form any coalition but will simply adopt their individually best reply strategies. This results into a non-cooperative equilibrium between the deviating coalition S and the nonmembers acting individually. The members of S play their joint best reply strategies to the individually best reply strategies of the nonmembers. The justification for the assumption that the nonmembers

---

[3] These have been studied and contrasted to each other in various externalities contexts by Scarf (1971), Starret (1972), and Maler (1989) and in the public goods context by Foley (1970), Roberts (1974), Moulin (1987), and Chander (1993) among others. It is well-known that because of the underlying minimax or maximin assumptions, the α - and β - characteristic functions lead to large cores. In fact, as noted by Maler (1989) and Ray and Vohra (1997), in some cases the α - and β - cores may include the whole set of Pareto optima.

act individually comes from the fact (see Chander and Tulkens (1997, fn.6, p.387)) that if the nonmembers form one or more non-singleton coalitions, then the payoff of S defined by the resulting Nash equilibrium between the coalitions would only be higher. The assumption that the nonmembers act individually is thus equivalent to granting S a certain degree of pessimism.[4] In other words, the assumption is not concerning which coalitions will form after the deviation, but rather which coalitions the deviating coalition S thinks will or will not form, and in fact S presumes that it is the worst possible coalition structure that will form.[5] The uncertainty regarding the coalition structure subsequent to a deviation is thus resolved by assuming that the deviating coalition presumes that coalition structure that is worst from its point of view. Chander and Tulkens (1997) then show that the so-defined γ-characteristic function implies stability of the grand coalition in at least the games that they consider.

The purpose of this paper is to reinterpret the γ-core and justify its assumption as well as its prediction that the efficient coalition structure is stable in terms of the coalition formation theory (see the excellent recent survey by Yi (2003) and the references therein). We assume that coalitions can freely merge or break apart and are farsighted (that is, it is the final and not the immediate payoffs that matter to the coalitions). We then show that subsequent to a deviation by a coalition, it is ex post optimal for the nonmembers to break apart into singletons, as is assumed in the definition of the γ-characteristic function, and the grand coalition is a stable coalition structure. This stands in contrast to some results in the coalition formation theory which often rule out stability of the efficient coalition structure. Thus, our analysis serves a dual purpose. On the one hand, it justifies the γ-core in terms of the coalition formation theory, and on the other, it clarifies that in order to rule out the stability of the efficient coalition structures one must

---

[4] This is pessimism of a different sort: it is not concerning the strategies that will be adopted by the non-members (as in the case of α - and β - characteristic functions), but about the coalition structures that will be formed.

[5] Chander and Tulkens (1995, 1997) consider games that imply positive externalities from coalition formation. For games that imply negative externalities from coalition formation, the corresponding assumption underlying the γ-characteristic function is that S assumes that the coalition $N \setminus S$ would form and its members would adopt the best response joint strategies.

either place some exogenous restrictions on the process of coalition formation or assume that the coalitions are not farsighted.

Though our analysis is applicable to a wide variety of situations, in order to be more concrete, we confine ourselves to the original economic model of Chander and Tulkens (1995, 1997). It is easily seen that similar results apply to the symmetric oligopoly model as well as to the public good economy (Ray and Vohra (1997) and Yi (1997)).

The contents of this paper are as follows. In Section 2, we state the model and the concept of $\gamma-$core. In Section 3, we consider a simple example that illustrates the γ-core and its relationship with the coalition formation theory. Section 4 presents the main results. Section 5 draws the conclusion.

2. The Model

Consider an economy consisting of n identical countries or agents. Let $N = \{1, 2, \dots, n\}$ denote the set of these agents. There are two kinds of commodities: a standard private good, whose quantities are denoted by $y,$ and an environmental good (in fact, a bad), whose quantities are denoted by $z.$ Suppose that the private good and the environmental good can be produced by the agents according to the following general rules:

$$y_i = g(e_i), i \in N, \text{ and } z = \sum_{i \in N} e_i, \tag{1}$$

where e_i is to be interpreted as the emissions of agent $i.$ We assume that $g'(e_i) > 0$ and $g''(e_i) < 0.$ The preferences of agent i are represented by the utility function:

$$u_i(y_i, z) = y_i - v(z), i \in N, \tag{2}$$

where $v'(z) > 0$ and $v''(z) \geq 0.$

We assume that there exists a finite e^0 such that $g'(e^0) < v'(e^0)$ and $g'(0) > nv'(e^0).$

This assumption rules out corner solutions and ensures that the emissions of each utility maximizing agent are strictly positive but not higher than e^0.

Let $(e_1^*, e_2^*, \ldots, e_n^*)$ be the Pareto efficient emissions.[6] Then, the first order conditions imply

$$g'(e_i^*) = nv'(z^*), \text{ and } e_i^* = e_j^*, \; i, j \in N. \tag{3}$$

Let $u_i^* = g(e_i^*) - v(z^*), \; i \in N,$ be the corresponding payoffs. Clearly $u_i^* = u_j^*,$ for $i, j \in N.$

Let $T_i = \{e_i : 0 \le e_i \le e^0\}, T = T_1 \times T_2 \times \cdots \times T_n,$ and $u = (u_1, u_2, \ldots, u_n).$ We consider the strategic form game $[N, T, u]$.

Let $(\bar{e}_1, \bar{e}_2, \ldots, \bar{e}_n)$ denote the Nash equilibrium of the game $[N, T, u].$ Then,

$$g'(\bar{e}_i) = v'(\sum_{j \in N} \bar{e}_j), i = 1, 2, \ldots, n. \tag{4}$$

Let $(\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_n)$ denote the payoffs corresponding to the Nash equilibrium, that is, $\bar{u}_i = g(\bar{e}_i) - v(\bar{z}), \bar{z} = \sum_{j \in N} \bar{e}_j$.

A *partition* of N is $P = (S_1, S_2, \ldots, S_m)$ such that $\bigcup_{j=1}^m S_j = N$ and for all $i \ne j, \; S_i \cap S_j = \Phi.$ Let n_j denote the cardinality of S_j, i.e., $n_j = |S_j|.$ Since the players are identical, we may denote coalition S_j interchangeably by n_j. Those partitions of N that consist of a possibly non-singleton coalition S followed by one or more coalitions of singletons are of particular interest. We denote such a partition simply by $(S, 1).$ The finest partition of N consisting of all singletons is denoted by $P^0 \equiv (1,1).$ We shall interchangeably refer to a partition $P = (S_1, S_2, \ldots, S_m)$ as the set of coalitions

---

[6] The Pareto efficient emission levels are unique (see Chander and Tulkens (1997)).

$S_1, S_2, \ldots, S_m$, i.e., as $\{S_1, S_2, \ldots, S_m\}$. A partition P is called a *coalition structure* and let $\wp$ be the *set of all coalition structures.* The idea of non-cooperative play across coalitions in a coalition structure is captured in the following definition.

Given a coalition structure $(S_1, S_2, \ldots, S_m) \in \wp$, the corresponding *coalitional equilibrium* $(\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_n)$ is defined as

$$(\tilde{e}_i)_{i \in S_j} = \text{argmax} \left(\sum_{i \in S_j} \left[g(e_i) - v\left(\sum_{i \in S_j} e_i + \sum_{k \in N \setminus S_j} \tilde{e}_k\right) \right] \right), \; j = 1, 2, \ldots, m. \tag{5}$$

Let $\tilde{u}_i = g(\tilde{e}_i) - v(\sum_{j \in N} \tilde{e}_j), i \in N,$ denote the corresponding payoffs of the players.

The existence and uniqueness of Nash equilibrium as well as coalitional equilibrium with respect to any given partition of N follow from standard arguments as the strategy sets are compact and convex and the payoff functions are concave (see e.g. Chander and Tulkens (1997)).

We are now well equipped to introduce the γ- characteristic function, to be denoted by w^{γ}. Let $S \subset N$ be some coalition and let $(\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_n)$ be the coalitional equilibrium corresponding to the coalition structure $(S, 1),$ that is

$$(\hat{e}_i)_{i \in S} = \text{argmax} \left(\sum_{i \in S} \left[g(e_i) - v\left(\sum_{i \in S} e_i + \sum_{j \in N \setminus S} \hat{e}_j\right) \right] \right)$$

and

$$\hat{e}_j = \text{argmax} \left[g(e_j) - v\left(\sum_{i \in N, i \neq j} \hat{e}_i + e_j\right) \right], \; j \in N \setminus S.$$

Let $\hat{u}_i = g(\hat{e}_i) - v(\sum_{j \in N} \hat{e}_j)$ and

$$w^{\gamma}(S) \equiv \sum_{i \in S} \left[g(\hat{e}_i) - v\left(\sum_{j \in N} \hat{e}_j\right) \right] = \sum_{i \in S} \hat{u}_i. \tag{6}$$

Let $(x_1, x_2, \ldots, x_n)$ be an imputation of the *n*-person game $[N, w^\gamma]$, that is $\sum_{i \in N} x_i = w^\gamma(N)$. Clearly $w^\gamma(N) = \sum_{i \in N} u_i^*$, where u_i^*'s are as defined in (3). Thus, $\sum_{i \in N} x_i = \sum_{i \in N} u_i^*$. However, this does not mean $x_i = u_i^*$. Thus, an imputation may require transfers among the members of the grand coalition.

Theorem (Chander and Tulkens (1997)): The game $[N, w^\gamma]$ has a nonempty core in the sense that there exists an imputation $(x_1, x_2, \ldots, x_n)$ such that $\sum_{i \in S} x_i > w^\gamma(S)$ for all $S \subset N, S \neq N.$

Though the original definition of the γ-core allows transfers among the members of a coalition, we will ignore this possibility here.[7] Since all players are identical, this does not imply an empty core.

Corollary 1: The imputation $(u_1^*, u_2^*, \ldots, u_n^*)$ assigning equal payoffs to all players belongs to the core of the game $[N, w^\gamma]$, that is, $w^\gamma(S) < \sum_{i \in S} u_i^*$ for all $S \subset N, S \neq N.$

Corollary 2: Let $(\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_n)$ denote the coalitional equilibrium corresponding to the coalition structure $(S, 1)$ with $S \neq N$ and let $(\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n)$ be the corresponding payoffs, that is, $\hat{u}_i = g(\hat{e}_i) - v(\sum_{j \in N} \hat{e}_j)$. Then, $\hat{u}_i < u_i^*$ for all $i \in S.$

Corollary 2 follows from fact that $\sum_{i \in S} \hat{u}_i = w^\gamma(S) < \sum_{i \in S} u_i^*$, and $u_i^* = u_j^*$, $\hat{u}_i = \hat{u}_j$ for all $i, j \in S.$

In the above definition of the characteristic function, the worth of a coalition S is

---

[7] The original theorem shows that if the players are not identical, then transfers among the players may be necessary for achieving a core imputation. Helms (2001) shows that this game is balanced.

determined by an equilibrium concept, namely that of a coalitional equilibrium corresponding to the coalition structure $(S,1)$. This means that a deviating coalition neither takes as given the strategies of its complement, as in the case of strong or coalition proof Nash equilibria, nor presumes that the complement would follow minimax or maximin strategies, as in the case of the α - and β - characteristic functions. The deviating coalition, however, presumes that the complement $N \setminus S$ would break up into singletons. In particular, this means that any individual player, that is $|S| = 1,$ expects that a deviation by him alone will be enough to precipitate the disintegration of the remaining coalition of $N-1$ players into singletons. This 'all-or-none' expectation is central to the definition and existence (non-emptiness) of the γ-core. [8] It is also at the heart of the analysis that follows, as we show that it is in fact a credible expectation.

Chander and Tulkens (1997) justify the all-or-none expectation by claiming that it is equivalent to granting the deviating coalition S a certain degree of pessimism in the sense that it presumes that coalition structure that is worst from its point of view. We now offer an additional justification in terms of the coalition formation theory. We show that if coalitions can freely merge or break apart and are farsighted, then subsequent to a deviation by a coalition, it is ex post optimal for the nonmembers to break apart into singletons.

3. An Example

Our approach is best introduced by an example. Consider an economy consisting of three identical agents, i.e., $N = \{1, 2, 3\}$. Let

$$y_i = e_i^{\frac{1}{2}},\ i \in N,\ z = \sum_{i \in N} e_i,$$

and

---

[8] There is some empirical evidence in support of the all-or-none expectation. For instance, the Comprehensive Test Ban Treaty on nuclear testing can come into force *only if all* the current and potential nuclear powers sign it. Another case in point is the Kyoto Protocol. After the US refusal to be a party to the protocol, will the rest of the countries implement or abandon it? See Tulkens (1998) for an interesting discussion of this issue.

$$u_i(y_i, z) = y_i - z + \frac{1}{4},\ i \in N, \tag{7}$$

where e denotes emissions, y denotes output and z pollution.

Let $T_i = \{e_i : e_i \geq 0\}, T = T_1 \times T_2 \times T_3,$ and $u = (u_1, u_2, u_3).$ We consider the strategic form game $[N, T, u]$.

Standard arguments show that the game $[N, T, u]$ has a unique Nash equilibrium which induces the following state of the economy

$$\overline{e}_i = \frac{1}{4};\ \overline{y}_i = \frac{1}{2},\ i \in N;\ \overline{z} = \frac{3}{4};\ \text{and}\ \overline{u}_i = 0,\ i \in N, \tag{8}$$

where $(\overline{e}_1, \overline{e}_2, \overline{e}_3) \in T$ are the Nash equilibrium strategies. It is easily seen that a Pareto efficient state of the economy is given by

$$e_i^* = \frac{1}{36},\ y_i^* = \frac{1}{6},\ i \in N;\ z^* = \frac{1}{12};\ \text{and}\ u_i^* = \frac{1}{3};\ i \in N, \tag{9}$$

and that the emission levels are the same for all Pareto efficient states. We claim that the strategies (e_1^*, e_2^*, e_3^*) belong to the γ-core. Since the players are identical, we need to consider only two types of deviations, namely: a deviation by a coalition of any two players, say $\{1,2\}$ and a deviation by a coalition of any single player, say $\{3\}$.

Define $(\tilde{e}_1, \tilde{e}_2, \tilde{e}_3)$ such that $\tilde{e}_1, \tilde{e}_2$ = argmax $(e_1^{\frac{1}{2}} + e_2^{\frac{1}{2}} - 2e_1 - 2e_2 - 2\tilde{e}_3)$ and $\tilde{e}_3$ = argmax $(e_3^{\frac{1}{2}} - \tilde{e}_1 - \tilde{e}_2 - e_3)$. Then,

$$\tilde{e}_1 = \tilde{e}_2 = \frac{1}{16},\ \tilde{e}_3 = \frac{1}{4},\ \tilde{u}_1 = \tilde{u}_2 = \frac{1}{8}\ \text{and}\ \tilde{u}_3 = \frac{3}{8}. \tag{10}$$

The strategies $(\tilde{e}_1, \tilde{e}_2, \tilde{e}_3)$ represent the Nash equilibrium between the coalitions $\{1,2\}$ and $\{3\}$. By comparing the payoffs of the coalition $\{1,2\}$ under the strategies

$(\tilde{e}_1, \tilde{e}_2, \tilde{e}_3)$ and (e_1^*, e_2^*, e_3^*), it is seen that the coalition {1, 2} will not gain from the deviation.[9]

We now consider deviation by $\{3\}$ which really brings into focus the all-or-none expectation. When $S = \{3\}$ deviates it presumes that $N \setminus S = \{1,2\}$ will break up into singletons and the resulting equilibrium will be the Nash equilibrium between $\{3\}, \{1\}$ and $\{2\}$, which leads to the same payoffs as in (8). From a comparison of the payoffs in (8) and (9), it follows that coalition $\{3\}$ will also not gain from its deviation. This shows that the strategies (e_1^*, e_2^*, e_3^*) are in the γ-core of the strategic form game $[N, T, u]$.

But why should the coalition $\{1,2\}$ break up into singletons when {3} deviates? The stability of the grand coalition depends crucially on the answer to this question, as {3} will gain from its deviation if {1, 2} did not break up (compare the payoffs of {3} in (10) and (9)) and hence would engage in deviation.

Let us consider first the argument against breaking up of {1, 2}: if {3} deviates and {1, 2} does not breakup into singletons, then the resulting equilibrium and the corresponding payoffs of its members are as given in (10), which, as seen from (8), are higher than what their payoffs would be if it were to break up and induce the coalition structure ({1}, {2}, {3}). The coalition {1, 2} therefore should not break up. This argument however assumes implicitly that either the coalition structure ({1}, {2}, {3}) which emerges after the coalition {1, 2} breaks apart is final or the coalition {1, 2} is myopic and is concerned only with its immediate payoff.

Ray and Vohra (1997) assume the coalitions to be farsighted, but preclude the possibility of coalition merging, that is, coalitions can only become finer and not coarser. Under such an assumption the coalition structure ({1}, {2}, {3}) is indeed final as further

---

[9] It is worth noting that alternatively the α - and β - strategies require $\tilde{e}_3 \to \infty$ and $\tilde{u}_1, \tilde{u}_2 \to -\infty$, even though player 3 has a dominant strategy $\tilde{e}_3 = 1/4$.

deviations are not permitted. Their analysis, therefore, implies stability of the coalition structure ({1, 2}, {3}) but not of the grand coalition {1, 2, 3}. The grand coalition is also not stable in terms of the coalition formation games considered by Carraro and Siniscalco (1993), who assume the coalitions to be myopic, that is, they are concerned only with their immediate pay-offs.[10] It is easily seen that under their assumption too the coalition structure ({1, 2}, {3}) is stable.

Let us now introduce the possibility of coalition merging (Chander (1999)) in the above story. This creates the possibility of further continuations after the coalition {1, 2} breaks apart and induces the coalition structure ({1}, {2}, {3}). Indeed, the coalitions {1}, {2}, and {3} may merge and form the grand coalition {1, 2, 3} which gives to each merging coalition a higher payoff (compare the payoffs in (8) and (9)). Contrary to the argument discussed above, it is ex post optimal for the farsighted coalition {1, 2} to break apart and induce the temporary coalition structure ({1}, {2}, {3}) if it thinks that that would lead to the formation of the grand coalition and therefore to payoffs to its members which are strictly higher than if it did not break apart (compare the payoffs of members of {1, 2} under (10) and (9)).

Formation of the grand coalition from the coalition structure ({1}, {2}, {3}) however may not seem to be the only possibility. There seem to be at least three other possibilities, namely ({1, 2}, {3}), ({1, 3}, {2}), and ({2, 3}, {1}). Which of these alternative coalition structures is likely to form from the coalition structure ({1}, {2}, {3})?

The coalition formation theory considers alternative approaches to deal with the issue of multiple continuations after a single deviation (see e.g. Greenberg (1990), Chwe (1994), Ray and Vohra (1997)): the initial deviating coalitions may evaluate the subsequent other deviations in optimistic or pessimistic ways. For example, in the definition of the von Neumann and Morgenstern abstract stable set, a coalition deviates as long as this deviation might lead to some final outcome that benefits its members. In contrast, in the definition of the largest consistent set (Chwe (1994)), a coalition deviates only if its

---

[10] See also d’Aspremont and Gabszewicz (1986) and Barrett (1994; 2003).

members benefit from all final outcomes that its deviation may lead to. The assumption underlying the γ-core is in the same vein, as it is equivalent to assuming that a coalition may engage in a deviation only if its members stand to strictly benefit from the outcome that gives the least payoffs from among all the final outcomes its deviation may lead to. Any other final outcomes, even if they strictly benefit the deviating coalition, are irrelevant and the possibility of just one final outcome which does not strictly benefit the members of the deviating coalition is enough to deter it from deviation.

We now consider an alternative approach and show that the grand coalition is, in fact, the only coalition structure that can form from the finest coalition structure ({1}, {2}, {3}). This eliminates the multiplicity of continuations after a deviation and the need for assigning any notions of optimism or pessimism to the deviating coalition in order to choose among them.

Note that the formation of coalition structures other than the grand coalition such as ({1, 2}, {3}) from the finest coalition structure ({1}, {2}, {3}) is possible only if players cannot commit to not form coalitions that do not include all players. If the players are unable to make such a commitment, the situation can develop into a war of attrition in which each player waits for the other two to form a coalition in the hope of free-riding on them. This situation is aptly described by a two- stage game that begins from the finest coalition structure $P^0 = (\{1\},\{2\},\{3\})$ as the status quo. In Stage 1, each player acting alone may propose either to form a coalition with other players: <u>C</u> or to free-ride, i.e., to not form a coalition with any other player: <u>NC</u>. All players who announce C form a coalition and the rest of the players, if any, remain singletons.[11] In Stage 2, the players choose their emission levels. The payoffs of the players are as defined in (7).

This game clearly has an equilibrium in which each player announces NC at Stage 1 and at Stage 2 chooses his emission level equal to that in the Nash equilibrium (see (8) above). This is an equilibrium because no individual player can improve his payoff by

---

[11] For the three agent case, restricting the choice of strategies to only C and NC does not rule out any coalition structure. For the *n* agent case, however, we allow more general strategies.

switching to C and choosing a different emission level, given the strategies of the other players. However, if the game results into this equilibrium, then everyone stands to gain from playing the game all over again. Thus, we consider repeated plays of the game. However, restricting this game to a finite number of repetitions is the same thing as restricting it to a single play. Therefore, we consider infinite repetitions.

We show that the infinitely repeated two-stage game has an efficient equilibrium with the grand coalition as the equilibrium coalition structure. In this equilibrium, each player's expectation is as follows: if everyone announces C at Stage 1, then at Stage 2 the emission levels and the payoffs will be as in (9). But if any player announces NC at Stage 1, then at Stage 2 coalitions will not be given effect and their members will choose their emission levels so as to maximize their individual payoffs. Therefore, the emission levels and the payoffs of all players will be the same as in the Nash equilibrium and the game will be repeated next period, as everyone stands to gain from it.

In other words, each player expects that if any player announces NC at Stage 1, then at Stage 2 coalitions will in effect break apart, as is assumed in the definition of the $\gamma-$characteristic function, and each player will choose his emission level so as to maximize his individual payoff. For this reason we call it the 'all-or-none' expectation. We show that the all-or-none expectation of the players is rational in the sense that it is ex post optimal for the coalitions to break apart, if any player chose NC at Stage 1. In other words, given this expectation of the players, it is credible for the players to commit to refrain from giving effect to any coalition that does not include all players.

Let w denote the equilibrium payoff or value of the repeated game to each player. Since the players are identical, this value is the same for all players. We solve for the equilibrium first with discounting, and then take the limit as the discount rate goes to zero. Let β denote the discount factor, i.e., $\beta = 1/(1+r)$ where r is the discount rate. The Stage 1 payoff matrix (below) is defined as follows: if all players choose C, the grand coalition is formed and each player gets $1/3$, as in (9). If any player chooses NC, then given the assumed all-or-none expectation, the emission level of each player will be the same as in the Nash equilibrium and the game will be repeated. The payoff to each player from this is w starting one period later which therefore has a discounted present value of βw in the current period.

Player 3

	C	NC
Both Player 1 and 2 choose C	$\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$	$\beta w, \beta w, \beta w$
Player 1 and/or 2 choose NC	$\beta w, \beta w, \beta w$	$\beta w, \beta w, \beta w$

In order to find a solution to this reduced Stage 1 game, we allow for mixed strategies. Let $p_1, p_2,$ and p_3 be the probabilities assigned by the three players to the C strategy. Then, in equilibrium, we must have

$$w = p_1 p_2 \frac{1}{3} + (1 - p_1 p_2)\beta w$$

$$w = p_1 p_2 \beta w + (1 - p_1 p_2)\beta w$$

Since these equations do not have a solution for any $\beta < 1,$ the game must have a Nash equilibrium in pure strategies. However, NC cannot be a Nash equilibrium strategy for any player, since that would imply the impossibility $w = \beta w.$ Therefore, C is the Nash equilibrium strategy for each player and the resulting equilibrium payoff is $w = \frac{1}{3}$. Since this implies $\beta w < \frac{1}{3} (= w)$ for $\beta < 1,$ C is actually a dominant strategy. It remains to be shown that the all-or-none expectation is credible. Suppose some player, say 3, tries to test it by choosing NC at Stage 1. Then a deviation from the all-or-none strategy by players 1 and 2 (that is, both choose their emissions as in (10) and not as in (8)) will get them each $1/8$. But if they stick to it, the game will be repeated and each of them will get $\beta w = \beta\, 1/3 > 1/8$ for β close enough to 1. This proves that the all-or-none strategy is ex post optimal and the resulting equilibrium is subgame-perfect.

Note that the argument here is not that players 1 and 2 may force player 3 to join the coalition by threatening that they will choose their emission levels to be equal to those in the Nash equilibrium (and thereby deny him the opportunity to free-ride) unless he joins them , but that such an action is ex post optimal for players 1 and 2. Therefore, this is a credible strategy for players 1 and 2 that cannot be ignored by player 3 when choosing independently his strategy at Stage 1.

Though we have found an efficient equilibrium of the repeated game, it is not unique. Like many other repeated games this one too has multiple equilibria. However, we show that the efficient equilibrium is the focal point (see Schelling (1960)) and therefore the most likely outcome.

The three coalition structures in which exactly two players choose C and the remaining player chooses NC can be obtained as equilibrium coalition structures by assuming a low discount factor β. This is so, in particular, if β is such that $\beta \frac{1}{3} < 1/8$.[12] However, most of coalition formation theory assumes zero discounting, i.e. $\beta = 1$. Furthermore, these equilibria are not symmetric and require coordination among players assumed to choose their strategies independently at Stage 1.[13] Since this is inconsistent with the non-

---

[12] This means that the players must be sufficiently impatient if these equilibria are to obtain, and a low discount factor has the same effect as restricting the game to a single play.

[13] See Farrell and Saloner (1988) and Crawford and Haller (1990) on why asymmetric pure-strategy equilibria may be unconvincing and inappropriate in coordination games. An alternative approach might be to assume that these equilibria are played randomly and all the three equilibria are equally likely, and take the average of the players' payoffs across these equilibria as their expected payoffs. However, since these equilibria are generally not efficient, the average payoffs are each below the payoffs in the efficient equilibrium. As seen from this example, the players' average payoffs of $\frac{1}{3}(\frac{3}{8}+\frac{1}{8}+\frac{1}{8})$ each are less than the players' payoffs of $\frac{1}{3}$ each under the efficient equilibrium.

cooperative nature of Stage 1, we must discard these equilibria in favor of symmetric equilibria. However, there are no symmetric equilibria in pure strategies other than the efficient equilibrium. We are thus left with symmetric mixed-strategy equilibria. There indeed exists one such equilibrium, but the players' payoffs under this equilibrium are equal (by definition) and generally below the payoffs corresponding to the efficient equilibrium, since the probability for the grand coalition to form is less than 1. The inefficiency of this equilibrium may then seem to imply that the equilibrium coalition structures may not necessarily be efficient. But this is not so. On the contrary, its inefficiency implies that the efficient equilibrium is the focal point. Since each player has full knowledge of the data concerning the game, each can independently find out that the efficient equilibrium offers higher payoffs to everyone than any other. Each player should then think that everyone else must have done the same calculations and conclude that the efficient equilibrium is to be favored by everyone. Therefore, each player should think that all will think that … each player will play the strategy corresponding to the efficient equilibrium.

4. The General Case

We now extend the results obtained in the example above to the general model. Corollary 2 to the Theorem plays a crucial role in this extension. We assume that each player can choose more generally from among the strategies consisting of NC and any number from 1 to n. The players who announce the same number form a coalition, but the players who choose NC remain singletons. Thus the players can induce any coalition structure of the form $(S_1, S_2, \ldots, S_m)$.

As in the example above, it is straightforward to see that given the all-or-none expectation, C is a dominant strategy for each player in the reduced Stage 1 game and the resulting equilibrium payoff is $w = u_i^*$. Therefore, it only remains to check that the all-or-none strategy is ex post optimal. This means that we must show that the players cannot gain by forming coalitions other than the grand coalition. We need the following lemma.

Lemma 3: Let $P = (S_1, S_2, \dots, S_m)$ be some coalition structure with $P \neq N$. Then the payoffs of the members of the largest coalition in the corresponding coalitional equilibrium are lower than their payoffs in the grand coalition, i.e., $\tilde{u}_i < u_i^*$ for all $i \in S_k$ such that $|S_k| \geq |S_j|$ for all j.

Proof of Lemma 3: Let $(\tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_n)$ be the coalitional equilibrium corresponding to the coalition structure $(S_1, S_2, \dots, S_m)$. Then the first order conditions imply

$$g'(\tilde{e}_i) = |S_j| \, v'(\sum_{k \in N} \tilde{e}_k), i \in S_j, \, j = 1,2,\dots,m. \tag{11}$$

Since g is strictly concave, $\tilde{e}_i < \tilde{e}_j$ if $i \in S_k$ and $j \in S_l$ with $|S_k| > |S_l|$. Let $\tilde{z} = \sum_{j \in N} \tilde{e}_j$ denote the total emissions corresponding to the coalitional equilibrium with respect to $(S_1, S_2, \dots, S_m)$. Then, $\tilde{u}_i \equiv g(\tilde{e}_i) - v(\tilde{z}) < \tilde{u}_j \equiv g(\tilde{e}_j) - v(\tilde{z})$ if $i \in S_k$ and $j \in S_l$ with $|S_k| > |S_l|$, since as shown $\tilde{e}_i < \tilde{e}_j$. Thus the payoffs of the members of larger coalitions are lower. Furthermore, by comparing (11) with the optimality condition (3), it follows that $\sum_{i \in N} \tilde{u}_i < \sum_{i \in N} u_i^*$ if $P \neq N$. Therefore, we must have $\tilde{u}_i < u_i^*$ for all $i \in S_k$ such that $|S_k| \geq |S_j|$ for all j.

Now suppose the players try to test the all-or-none strategy by forming coalitions $(S_1, S_2, \dots, S_m) \neq N$. Assume without loss of generality that $|S_1| \leq |S_2| \leq \cdots \leq |S_m|$. Consider the finite sequence of coalition structures $P^m \equiv (S_1, S_2, \dots, S_m)$, $P^{m-1} \equiv (S_1, S_2, \dots, S_{m-1}, 1)$, $P^{m-2} \equiv (S_1, S_2, \dots, S_{m-2}, 1)$, …, $P^1 \equiv (S_1, 1)$. This sequence of coalition structures is obtained if each time the largest coalition in the coalition structure breaks up into singletons starting from the largest coalition S_m in P^m. Let $\tilde{u}_i^k, i \in S_k, k = 1,2,\dots,m$ be the corresponding sequence of payoffs, that is, $\tilde{u}_i^k$, $i \in S_k$, are the payoffs of the members of the largest coalition in the coalition structure $(S_1, S_2, \ldots, S_k, 1)$. Then, by Lemma 3, $\tilde{u}_i^k < u_i^*$ for all $i \in S_k, k = 1,2,\ldots, m.$ Therefore,

$$\tilde{u}_i^k < \beta u_i^* \text{ for all } i \in S_k, k = 1,2,\ldots, m, \tag{12}$$

for β sufficiently close to 1.

Begin with the partition $P^1 = (S_1, 1),\ S_1 \neq N.$ Suppose some players deviate from the all-or-none strategy by forming the coalition S_1 when the rest of the players have chosen NC. Then their payoffs will be $\tilde{u}_i^1, i \in S_1.$ But if they stick to the all-or-none strategy (that is break apart) the game will be repeated and each of them will get a payoff of $w = u_i^*$ starting one period later which has a discounted present value of βu_i^* in the current period. Since as seen from (12), $\beta u_i^* > \tilde{u}_i^1,\ i \in S_1,$ for β sufficiently close to 1, it is optimal for the members of S_1 to break apart. Consider next the partition $P^2 = (S_1, S_2, 1).$ Suppose some players deviate from the all-or-none strategy by forming coalitions S_1 and S_2 when the rest of the players, if any, have chosen NC. Assume without loss of generality that $|S_2| \geq |S_1|.$ Then the payoffs of members of S_2 will be $\tilde{u}_i^2, i \in S_2.$ But if the members of S_2 follow the all-or-none strategy and break apart, then, as shown above, the members of S_1 will also break apart and the game will be repeated and their payoffs will be equivalent to $\beta u_i^* > \tilde{u}_i^2,\ i \in S_2.$ Hence, players in $|S_2|$ will break apart. But then the players in $|S_1|$ will also break apart, as $\beta u_i^* > \tilde{u}_i^1, i \in S_1.$ This proves that it is optimal for the players to stick to the all-or-none strategy and not form separate coalitions at Stage 2. By induction, it follows that the players cannot gain by forming separate coalitions $(S_1, S_2, \ldots, S_m).$ This proves that the all-or-none strategy is ex post optimal and the resulting equilibrium is subgame-perfect.

The process of disintegration of the coalitions in the sequence described above must continue until the finest coalition structure P^0 is reached because every coalition structure other than the grand coalition has a largest coalition whose members have lower payoffs compared to their payoffs in the grand coalition. Disintegration of coalitions imposes negative externalities on other coalitions and weakens their incentives to break away from the grand coalition. We confirm this intuition in the following proposition.

Given $P = (S_1, S_2)$, let $P' = (S_1, 1)$ denote the coalition structure which is obtained if S_2 breaks up into singletons. Since the payoff of each player in a coalitional equilibrium depends on the entire coalition structure, let $u_i : \wp \rightarrow R$ denote i's payoff.

Proposition 4: Let $P = (S_1, S_2)$ be some coalition structure such that $|S_2| > 1$ and let $P' = (S_1, 1)$. Then $u_i(P') < u_i(P)$ for all $i \in S_1$.

Proof of Proposition 4: Let $(\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_n)$ and $(e'_1, e'_2, \ldots, e'_n)$ be the coalitional equilibrium strategies corresponding to P and P', respectively. Let $\tilde{z} = \sum_{i \in N} \tilde{e}_i$ and $z' = \sum_{i \in N} e'_i$. We claim that $z' > \tilde{z}$. Suppose not, i.e., let $z' \leq \tilde{z}$. Then, from (11), given strict concavity of g and convexity of v, $e'_i \geq \tilde{e}_i$ for each $i \in S_1$, and $e'_i > \tilde{e}_i$ for each $i \in S_2$, since $|S_2| > 1$. But this contradicts our supposition that $z' \leq \tilde{z}$. Hence, $z' > \tilde{z}$ and $e'_i \leq \tilde{e}_i$ for each $i \in S_1$ which implies $u_i(P') < u_i(P)$ for $i \in S_1$. This completes the proof.

Conversely, this proposition means that formation of a coalition by some players results into positive externalities for the remaining players and provides them stronger incentives to free ride. Intuitively, when a coalition forms its members internalize the effect of their emissions on each other resulting into lower emissions by them which benefits the other players as well. Disintegration of coalitions has the opposite effect. Examples are easily constructed in which $u_i(P) > u_i^*$ for all $i \in S_1$ but $u_i(P') < u_i^*$ for all $i \in S_1$ illustrating that disintegration of coalition S_2 may reverse the incentives of coalition S_1 to break away from the grand coalition.

We have shown that the grand coalition is the *only* coalition structure that can form from the finest coalition structure. If the coalition formation process reaches or begins from the finest coalition structure, then it will *end* with the formation of the grand coalition as the final coalition structure. This result is the key to the stability of the grand coalition as it rules out cycles and effectively eliminates the multiplicity of continuations after a deviation by a coalition.

Our analysis may seem to apply only to the case in which the coalition formation process starts from the finest coalition structure. But this is not so. Let $P = (S_1, S_2, \ldots, S_m)$, $P \neq P^0, N,$ be some coalition structure. Assume without loss of generality that $|S_1| \leq |S_2| \leq \cdots \leq |S_m|$. Consider the sequence of coalition structures $P^m = P = (S_1, S_2, \ldots, S_m)$, $P^{m-1} = (S_1, S_2, \ldots, S_{m-1}, 1)$, $P^{m-2} = (S_1, S_2, \ldots, S_{m-2}, 1)$, $\ldots$, $P^1 = (S_1, 1)$, $P^0 = (1,1)$, $N.$ This sequence is obtained if starting from the largest coalition S_m in the original coalition structure P^m, the largest coalition in each subsequent coalition structure also breaks apart until the finest coalition structure P^0 is reached, after which, as shown, the players will regroup and form the coalition $N.$ Since $u_i^* > u_i(P^k)$ for all $i \in S_k$, $k = 0,1,2,\ldots, m,$ by Lemma 3, each largest coalition in the sequence stands to gain ultimately by breaking apart. Thus the farsighted coalitions will continue to disintegrate until the finest coalition structure P^0 is reached and the grand coalition is formed.

We now formally define what we mean by a stable coalition structure. A coalition structure $P = (S_1, S_2, \ldots, S_m)$ is *stable* if no coalition can strictly improve its payoff by engaging in a deviation.

Proposition 5: The grand coalition N is stable.

Proof of Proposition 5: Suppose a coalition S breaks away from N. There are two possible cases: either $|S| \leq |N \setminus S|$ or $|S| > |N \setminus S|$. Let $P = (S, N \setminus S)$ and $P' = (S,1)$.

Consider first $|S| \leq |N \setminus S|$. Then, in view of Lemma 3, $u_i(P) < u_i^*$ for all $i \in N \setminus S$ and, by Corollary 2 to the Theorem, $u_i(P') < u_i^*$ for all $i \in S$. This means that, given the coalition structure $P = (S, N \setminus S)$, the members of $N \setminus S$ stand to gain ultimately if they break apart, since that would induce the members of S also to break apart and lead back to the formation of the grand coalition. This proves that coalition S cannot strictly improve its payoff by breaking away from N.

If $|S| > |N \setminus S|$, then by interchanging S and $N \setminus S$ and applying the same argument as in the preceding paragraph, S will break apart first, followed by the breaking up of $N \setminus S$, and finally back to the formation of N. This again means that S cannot strictly improve its payoff by breaking away from N.

The above leaves out the possibility that rather than breaking up into singletons, $N \setminus S$ or S may break up into two or more non –singleton coalitions. However, there is no loss of generality in ignoring this possibility as that would only lead to intermediate coalition structures of the form that we have already shown will also lead back to the formation of the grand coalition N. This completes the proof.

5. Conclusions

We have shown that the players can credibly commit to refrain from forming coalitions that do not include all players and this implies stability of the grand coalition. Our analysis shows that in order to obtain stability of coalition structures other than the grand coalition or the efficient coalition structure one must either invoke the assumption of myopia or place some exogenous restrictions on the process of coalition formation.[14]

The purpose of this paper is not to review the coalition formation theory, but to justify, in terms of this theory, the γ- core and its prediction that the efficient coalition structure is stable. Accordingly, as in the coalition formation theory, we have assumed the agents to be identical. However, our results do not depend on this assumption. As shown originally, the γ-core is non-empty even if the agents are not identical and a result similar to Corollary 2 holds with appropriate transfers among the agents. Lemma 3 does not hold as such because when the players are not identical their payoffs neither depend on the size of the coalitions nor are comparable across coalitions. However, in each coalition structure $P = (S_1, S_2, \ldots, S_m),\ P \neq N,$ there exists a coalition S_i (which is not necessarily the largest) whose members have lower payoffs compared to their payoffs in the grand coalition. These generalizations of Corollary 2 and Lemma 3 are sufficient to extend Propositions 4 and 5 to the case of asymmetric agents.

Though we have confined ourselves to a particular model, our analysis is applicable to other settings that have been considered in the literature such as the symmetric oligopoly model and the public good economy. This is so because results similar to Corollary 2 and Lemma 3 that drive our analysis hold in these models as well.

We have restricted our analysis to a class of games that imply positive externalities from coalition formation. For games with negative externalities, the corresponding assumption underlying the γ-characteristic function is that the deviating coalition S assumes that the residual coalition $N \setminus S$ will form and its members will adopt the best response *joint* strategies. We however do not pursue this here as this requires a different model in which coalition formation imposes negative externalities.[15]

---

[14] As noted earlier, efficiency may also not obtain if the agents are very impatient.

[15] See Belleflamme (2000) for a model with not only negative externalities but also asymmetric agents.

References

1. Belleflamme, P. (2000), “Stable coalition structures with open membership and asymmetric firms”, *Games and Economic Behavior*, 30, 1-21.
2. Barrett, S. (1994), "Self-enforcing international environmental agreements", *Oxford Economic Papers*, 46, 804-78.
3. Barrett, S. (2003), *Environment and Statecraft*, Oxford University Press, Chapter 7.
4. Carraro, C. and D. Siniscalco (1993), "Strategies for the international protection of the environment", *Journal of Public Economics*, 52, 309-28.
5. Chander, P. (1993), "Dynamic procedures and incentives in public good economies", *Econometrica*, 61, 1341-54.
6. Chander, P. and H. Tulkens (1995), "A core-theoretic solution for the design of cooperative agreements on transfrontier pollution", *International Tax and Public Finance*, 2, 279-93.
7. Chander, P. and H. Tulkens (1997), "The core of an economy with multilateral environmental externalities", *International Journal of Game Theory*, 26, 379-401.
8. Chander, P. (1999), "International treaties on global pollution: a dynamic time-path analysis", in G. Ranis and L.K. Raut (eds.), *Trade, Growth and Development: essays in honor of T.N. Srinivasan*, Elsevier Science, Amsterdam.
9. Chwe, M. S.-Y. (1994), “Farsighted coalitional stability”, *Journal of Economic Theory*, 63, 299-325.
10. Crawford, V. and H. Haller (1990), “Learning how to cooperate”, *Econometrica*, 58, 571-595.
11. D'Aspremont, C. and J.J. Gabszewicz (1986), "On the stability of collusion", in J.E. Stiglitz and G.F. Mathewson (eds), *New Developments in the Analysis of Market Structure*, pp. 243-64, The MIT Press, Cambridge (MA).
12. Farrell, J. and G. Saloner (1988), “Coordination through committees and markets”, *Rand Journal of Economics*, 19, 235-252.
13. Foley, D. (1970), "Lindahl solution and the core of an economy with public goods", *Econometrica*, 38, 66-72.
14. Greenberg, J. (1990), *The Theory of Social Situations*, Cambridge University Press,

Cambridge.

15. Hart, S. and M. Kurz (1983), “Endogenous formation of coalitions”, *Econometrica,* 51, 1047-64.
16. Helms, Carsten (2001), “On the existence of a cooperative solution for a coalitional game with externalities”, *International Journal of Game Theory,* 30, 141-147.
17. Mäler, K.-G. (1989), “The acid rain game”, in H. Folmer and E. van Ierland (eds), *Valuation Methods and Policy Making in Environmental Economics*, 231-52, Elsevier, Amsterdam.
18. Moulin, H. (1987), “Egalitarian-equivalent cost sharing of a public good”, *Econometrica*, 55, 963-76.
19. Ray, D. and R. Vohra (1997), “Equilibrium binding agreements”, *Journal of Economic Theory*, 73, 30-78.
20. Roberts, D.J. (1974), “The Lindahl solution for economies with public goods”, *Journal of Public Economics,* 3, 23-42.
21. Scarf, H. (1971), "On the existence of a cooperative solution for a general class of *N*-person games, *Journal of Economic Theory*, 3, 169-81.
22. Schelling, T.C. (1960), *The Strategy of Conflict*, Harvard University Press, Cambridge, MA.
23. Starrett, D. (1972), "A note on externalities and the core", *Econometrica*, 41, 179-83.
24. Tulkens, H. (1998), “Cooperation versus free-riding in international environmental affairs: two approaches” in N. Hanley and H. Folmer (eds.), *Game Theory And The Environment*, 30-44, Edward Elgar, UK.
25. Yi, S. (1997), "Stable coalition structures with externalities", *Games and Economic Behavior,* 20, 201-37.
26. Yi, S. (2003), “Endogenous formation of economic coalitions: a survey of partition function approach” in C. Carraro (ed.), *The Endogenous Formation of Economic Coalitions,* 80 -127, Edward Elgar, UK.