Int J Game Theory (2007) 35:539–556
DOI 10.1007/s00182-006-0067-9

ORIGINAL PAPER

The gamma-core and coalition formation

Parkash Chander

Accepted: 4 December 2006 / Published online: 24 January 2007

Abstract This paper reinterprets the γ-core (Chander and Tulkens in Int Tax Pub Financ 2:279–293, 1995; in Int J Game Theory 26:379–401, 1997) and justifies it as well as its prediction that the efficient coalition structure is stable in terms of the coalition formation theory. The problem of coalition formation is formulated as an infinitely repeated game in which the players must choose whether to cooperate or not. It is shown that a certain equilibrium of this game corresponds to the γ-core assumption that when a coalition forms the remaining players form singletons, and that the grand coalition is an equilibrium coalition structure.

JEL Classification numbers C71, C72, D62

Keywords Core · Strategic games · Nash equilibrium · 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 remaining players. If there are no externalities, i.e., if the payoffs to the members of a coa-

An earlier version of this paper was presented at the conference on Game Theory and Its Applications held in Mumbai in 2003 and was subsequently circulated as CORE Discussion Paper 2003/46.

P. Chander (✉)
National University of Singapore, 1 Arts Link, AS 2, Level 6, Singapore 117570, Singapore
e-mail: ecsparka@nus.edu.sg

lition do not depend on the actions of the non-members, then the worth can be defined without specifying the actions of the non-members. But if externalities are present, then in order to calculate the worth of a coalition one must also predict the actions of the non-members. This has been a debated issue for long, and the standard recipe for this until recently has been a minimax argument that leads to the concepts of α- and β-characteristic functions.[1] However, the minimax argument involves the assumption that if a coalition deviates, it fears the worst in the sense that it presumes that the outside players will act so as to minimize its payoff without concern for their own payoffs. Thus, by their very construction, the α- and β-characteristic functions assume away the interesting strategic interactions between the deviating coalition and the outside players.

However, the γ- characteristic function introduced recently by Chander and Tulkens (1995, 1997) is based on a more plausible assumption concerning the behavior of the non-members in that it is assumed that when a coalition deviates it presumes that the outside players will adopt their individually best reply strategies. This results into a non-cooperative equilibrium between the deviating coalition S and the other players acting individually, in which the members of S play their joint best reply strategies to the individually best reply strategies of the remaining players. Though, unlike the α- and β-characteristic functions, this brings into play the interesting strategic interactions between the deviating coalition and the outside players, it may appear arbitrary to assume that the outside players act as individuals and do not cooperate by forming non-trivial coalitions.[2] Until now this assumption has been justified on the grounds that if the outside players form non-trivial coalitions, then the payoff of S, defined by the resulting Nash equilibrium between the coalitions, will be generally higher. The assumption is thus equivalent to granting the deviating coalition S a certain degree of pessimism in the sense that it presumes the emergence of a coalition structure that is the worst from its point of view.[3] Though economically meaningful, this could be a very unreasonable forecast of what coalitions the outside players will actually form, and the deviating coalition may conceivably do better than simply presume the emergence of the worst coalition structure. It might be able to predict more accurately the coalition structure that would form subsequent to its deviation and not just presume that the complement would break apart into singletons.

This thinking naturally calls for a non-cooperative way of looking at the situation. The endogenous coalition formation theory [see e.g. Bloch (1996),

[1] These have been studied in various externalities contexts by Scarf (1971), Starrett (1972), and Mäler (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 assumption, the α- and β-characteristic functions lead to large cores. In fact, as noted by Ray and Vohra (1997), in some cases the α- and β-cores may include the entire set of Pareto optima.

[2] The γ-characteristic function is thus a restriction of the partition function (Thrall and Lucas 1963) in which for each coalition the rest of the coalition structure consists of singletons only.

[3] This is pessimism of a different sort: it is not concerning the strategies that will be adopted by the outside players (as in the case of α- and β-characteristic functions), but about the coalitions that will be formed by them subsequent to the deviation.

Yi (1997), and Ray and Vohra (1997, 1999, 2001)], which is concerned with such issues, indeed uses a non-cooperative approach. In particular, it characterizes equilibrium coalition structures in the context of strategic games with transferable utility in which it is efficient for the grand coalition of all agents to form and choose the strategy profile that maximizes the joint surplus.[4] It explores whether the agents will have an incentive to cooperate and form a coalition. The issue is not whether the coalition is profitable and how the profit may be shared among the coalition members, but whether the agents will have the incentives to form a coalition at all.[5] When coalitions generate significant positive externalities and an agent can derive benefit from a coalition's activity without joining it, the agent may have an incentive to remain apart, i.e., to free ride.

In this paper, we reinterpret the γ-core and justify the said assumption as well as its prediction that the efficient coalition structure is stable in terms of the coalition formation theory. We formulate the problem of coalition formation as an infinitely repeated game in which the players must choose whether to cooperate or not. It is shown that a certain equilibrium of this game corresponds to the γ-core assumption that when a coalition forms the remaining players form singletons, and that the grand coalition is an equilibrium coalition structure. Thus, our analysis reconciles a 'core' concept in the cooperative game theory with an equilibrium (coalition structure) concept in the non-cooperative game theory.

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 global pollution considered by Chander and Tulkens (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, 2001 and Yi 1997).

The contents of this paper are as follows. Section 2 describes the model and the concept of γ-core. Section 3 considers a simple example that illustrates the said γ-core assumption and its relationship with the coalition formation theory and formulates the problem of coalition formation as an infinitely repeated game. Section 4 presents the main results. Section 5 draws the conclusion.

2 The model

Consider an economy consisting of n identical agents. Let $N = \{1, 2, \ldots, n\}$ denote the set of these agents. There are two kinds of commodities: a standard private good, whose quantities are denoted by $y_1, y_2, \ldots, y_n$, 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 production technology and the transfer function, respectively,

[4] Coalition structures other than the grand coalition thus imply inefficiency.

[5] Of course, the latter depends on the way in which the coalition profit is shared among coalition members.

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

where e_i and y_i denote agent i's emissions and output, respectively, and z the total pollution. Assume that $g'(e_i) > 0$ and $g''(e_i) < 0$. Suppose that the preferences of agent i are represented by the utility function

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

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

Assume that there exists a finite e^0 such that $g'(e^0) < v'(e^0)$ and $g'(0) > v'(ne^0)$. 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

$$g'(e_i^*) = nv'(z^*), \tag{3}$$

and consequently $e_i^* = e_j^*, i, j \in N$. Let

$$u_i^* = g(e_i^*) - v(z^*), \quad i \in N, \tag{4}$$

be the corresponding payoffs. Clearly, $u_i^* = u_j^*$, for $i, j \in N$.

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

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

Let $(\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_n)$ be 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 $i \neq j, \ S_i \cap S_j = \phi$. 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 shall 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 $S_1, S_2, \ldots, S_m$, i.e., $\{S_1, S_2, \ldots, S_m\}$.

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.

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

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

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

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 a 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 in the strategies of all players (see e.g. Chander and Tulkens 1997).

2.1 The γ-characteristic function

The γ-characteristic function will 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,

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

and

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

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 characteristic function form 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 equation (4). Thus, $\sum_{i\in N} x_i = \sum_{i\in N} u_i^*$. This, however, does not mean $x_i = u_i^*$ and an imputation may require transfers among the members of the grand coalition.

Theorem 1 (Chander and Tulkens 1997) *The game* $[N, w^\gamma]$ *has a non-empty 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 players, we will ignore this possibility here.[7] Since all players are identical, ruling out transfers does not imply an empty core.

Corollary 2 *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 3 *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 3 follows from the 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 determined by an equilibrium concept, namely that of a coalitional equilibrium corresponding to the coalition structure $(S, 1)$. This means that a deviating coalition does not presume that the complement would follow minimax strategies, as in the case of the α- and β-characteristic functions. However, it presumes that the complement $N\backslash S$ would break apart into singletons. In particular, each individual player, that is $|S| = 1$, expects that a deviation by him *alone* will be sufficient to precipitate 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; and we show that this expectation is in fact rational.

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 the deviating coalition presumes the emergence of a coalition structure that is the worst from its point of view. We now offer an additional justification in terms of the endogenous coalition formation theory. We show that a certain equilibrium of an infinitely repeated game in which the players choose whether or not to cooperate corresponds to the γ-core assumption that when a coalition forms the remaining players form singletons.

[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. Helm (2001) shows that the game is generally balanced.

[8] There is some empirical evidence in support of this 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.

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

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

and

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

where $e_1, e_2, \ldots, e_n$ and $y_1, y_2, \ldots, y_n$ denote the emissions and outputs, respectively, and z the total 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 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

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

where $(\bar{e}_1, \bar{e}_2, \bar{e}_3) \in T$ are the Nash equilibrium strategies. As is easily seen, a Pareto efficient state of the economy, which also maximizes the payoff of the grand coalition, is given by

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

and that the emission levels are the same in all Pareto efficient states.

3.1 The γ-core and its rationale

In order to illustrate the assumptions underlying the γ-core, we check whether the imputation (u_1^*, u_2^*, u_3^*) belongs 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 $\tilde{e}_1, \tilde{e}_2 = \text{argmax}\,(e_1^{\frac{1}{2}} + e_2^{\frac{1}{2}} - 2e_1 - 2e_2 - 2\tilde{e}_3)$ and $\tilde{e}_3 = \text{argmax}\,(e_3^{\frac{1}{2}} - \tilde{e}_1 - \tilde{e}_2 - e_3)$. Then,

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

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 deviating 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 coalition {1, 2} will not gain if it engages in a deviation, since $\tilde{u}_1 + \tilde{u}_2 < u_1^* + u_2^*$.[9]

Now consider deviation by {3}, which brings into play the said assumption underlying the γ-characteristic function. When $S = \{3\}$ deviates it presumes that $N \backslash S = \{1, 2\}$ will break up into singletons and thus the resulting equilibrium will be the Nash equilibrium between {3}, {1} and {2}, which will lead 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 proves that the imputation (u_1^*, u_2^*, u_3^*) belongs to the γ-core of the game.

Turning now to the rationale, why should the coalition {1, 2} break apart into singletons when {3} deviates? The stability of the grand coalition depends crucially on the answer to this question, as {3} would gain from its deviation if coalition {1, 2} does not break apart [compare the payoffs of {3} in (10) and (9)]; and it would thus engage in deviation.

Consider first the argument against breaking up of {1, 2}: if {3} deviates and {1, 2} does not break apart, then the resulting equilibrium and the corresponding payoffs of its members will be as in (10), which, as seen from (8), are higher than what their payoffs will be if they were to break apart and induce the coalition structure ({1}, {2}, {3}). The coalition {1, 2} therefore should not break apart. However, this argument assumes 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 the immediate payoffs to its members. Let us elaborate on these two points.

As to the former, in Ray and Vohra (1997) coalitions are assumed to be farsighted, but they can only become finer and not coarser. The coalition structure ({1}, {2}, {3}) is then indeed final, since further deviations are not possible, i.e., coalitions cannot become more finer. This implies stability of the coalition structure ({1, 2}, {3}) and not that of the grand coalition {1, 2, 3}. As to the latter point, 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, the coalitions are concerned only with their immediate pay-offs.[10] Clearly, under this assumption too the coalition structure ({1, 2}, {3}) is stable.

Now suppose that the coalitions are not only farsighted but can also merge.[11] This creates the possibility of further continuations after the coalition {1, 2} breaks apart and induces the coalition structure ({1}, {2}, {3}): for example,

[9] It is worth noting that the minimax assumption underlying the α- and β-characteristic functions requires player 3 to choose strategy $\tilde{e}_3 = \infty$ even though player 3 has a dominant strategy $\tilde{e}_3 = 1/4$. This is clearly absurd.

[10] Also see D'Aspremont and Gabszewicz (1986) and Barrett (1994).

[11] Diamantoudi and Xue (2002) show that allowing coalition merging alone is not sufficient to rule out the inefficient equilibrium coalition structures obtained under the Ray and Vohra (1997) assumption. As shown by Mauleon and Vannetelbosch (2004), a further refinement of their solution concept is necessary.

formation of the grand coalition {1, 2, 3} which will give each merging coalition a higher payoff [compare the payoffs of {1}, {2}, and {3} under (8) and (9)].

Thus, contrary to the two arguments just discussed, it might be ex post optimal for the farsighted coalition {1, 2} to break apart and induce the temporary coalition structure ({1}, {2}, {3}), if it believes 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)].

However, besides the grand coalition there are three other possibilities, namely, the inefficient coalition structures, ({1, 2}, {3}), ({1, 3}, {2}), and ({2, 3}, {1}) that may form from the temporary coalition structure ({1}, {2}, {3}). In each of these inefficient coalition structures exactly two players cooperate and the remaining player gets to free ride.

3.2 Coalition formation as a non-cooperative game

The question thus arises as to which coalitions will form from the finest coalition structure ({1}, {2}, {3}). This is in some sense a more basic question than the question of which coalition structures are stable. But as just argued the stability of the grand coalition depends crucially on the answer to this question. We thus resolve it first.

The situation is aptly dealt with in terms of a two-stage non-cooperative game that begins from the finest coalition structure $P^0 = (\{1\}, \{2\}, \{3\})$ as the status quo. In Stage 1, each player acting independently may propose to cooperate with other players: C or to free-ride, i.e., to not cooperate with any other player: NC. All those players who announce C form a coalition and the rest of the players, if any, form singletons.[12] In Stage 2, the players choose their actions, i.e., emission levels. The payoffs of the players are defined as in (7). Which coalition structures can one expect as equilibria of this game?

The game clearly has three pure-strategy asymmetric equilibria in which exactly two players cooperate and the remaining player gets to free ride.[13] However, one faces the question of how to choose among these equilibria. In a non-cooperative environment with identical players, there is no way to coordinate which player "gets" to free ride and which two are the ones who "have" to cooperate. Therefore, we ignore these asymmetric equilibria, and consider only symmetric mixed-strategy equilibria in which the outcome is left to chance and everyone has an equal chance to be the free rider. Remember that not only are the players identical but the game also starts from the finest coalition structure, which is symmetric.[14]

[12] In the three agent case, restricting the choice of strategies to C and NC does not rule out a priori the formation of any coalition structure. In the n agent case, however, we consider more general strategies.

[13] The number of such equilibria depends on the number of players and can be very large.

[14] In this, we closely follow Holmstrom and Nalebuff (1992) who use a similar argument in favor of choosing symmetric mixed strategy equilibria in a context that also involves free riding.

Note first that the game has a symmetric equilibrium in pure strategies in which each player announces NC in Stage 1 and chooses the same action in Stage 2 as in the Nash equilibrium. This is an equilibrium because no individual player can improve his payoff by switching to C and choosing a different action, given the actions of the other players. But if the game results into this equilibrium, then everyone stands to gain from playing the game all over again. Thus, we must consider repeated plays of the game. As is easily seen, restricting the game to a finite number of repetitions is the same thing as restricting it to a single play. Therefore, we consider infinite repetitions.

3.3 The γ-core imputation as an equilibrium of a repeated game

We show that in the infinitely repeated two-stage game (a) each player can credibly commit not to form a coalition unless it includes all players, and (b) that, as a result, the grand coalition is an equilibrium outcome. This equilibrium is supported by each player's belief that if everyone announces C in Stage 1, then the actions and the payoffs in Stage 2 will be the same as in (9), i.e. $u_i^* = 1/3$ for all i, but if any player announces NC in Stage 1, then no coalition will be given effect and the players will choose their actions in Stage 2 so as to maximize their individual payoffs: the actions and the payoffs of all players will be thus the same as in the Nash equilibrium, i.e. $\bar{u}_i = 0$ for all i, and the game will be repeated next period, as everyone stands to gain from it.

We first establish (b) assuming (a), and then establish (a). Let w denote the equilibrium payoff or value of the repeated game to each player. This value is same for each player, since the players are identical. In order to obtain a sharper characterization, we solve for the equilibrium of the repeated game 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 players' beliefs, all players choose the same emission levels as in the Nash equilibrium, and the game is repeated next period. The payoff to each player from this will be w starting one period later, which has a discounted present value of βw in the current period.

	Player 3: C	Player 3: 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	$\beta w, \beta w, \beta w$	$\beta w, \beta w, \beta w$

We now find a solution to this reduced Stage 1 game. Let p_1, p_2, p_3 be the probabilities assigned by the three players to the C strategy. Then, in equilibrium, each player, say 3, should be indifferent between choosing the strategies C and NC. Thus, we must have

$$\begin{aligned} w &= p_1 p_2 \tfrac{1}{3} + (1 - p_1 p_2)\beta w \\ w &= p_1 p_2 \beta w + (1 - p_1 p_2)\beta w \end{aligned}$$

Since these equations do not have a solution for $\beta < 1$, the game must have a Nash equilibrium in pure strategies. However, NC cannot be a Nash equilibrium strategy, since that would imply the impossibility $w = \beta w$. Hence, C is the Nash equilibrium strategy for each player and the resulting equilibrium payoff is $w = 1/3$. Since this implies $\beta w < \frac{1}{3} (= w)$ for $\beta < 1$, C is actually a dominant strategy if $\beta < 1$.

Note that we obtain this *symmetric* equilibrium without applying the restriction that the equilibrium be symmetric. We will return to this point in the conclusion to the paper.

Coming to (a) now, it remains to be shown that given the players' beliefs the all-or-none strategy is ex post optimal. Suppose some player, say 3, tries to test it by choosing NC in Stage 1, then a deviation from the all-or-none strategy by players 1 and 2 [that is, if they give effect to the coalition and choose their emissions as in (10) and not as in (8)] will get them each $1/8$. But if they stick to the all-or-none strategy, the game will be repeated and each of them will get $\beta w = \beta 1/3 > 1/8$ for β close enough to 1.

The argument is not that players 1 and 2 can force player 3 to join the coalition by threatening that they will choose the same emission levels as in the Nash equilibrium (and thereby deny him the opportunity to free-ride) unless he joins the coalition, but that given the players' beliefs such an action is ex post optimal for players 1 and 2. Hence, this is a credible strategy for players 1 and 2 that player 3 cannot ignore when choosing his strategy in Stage 1.

We have thus found an efficient symmetric equilibrium of the repeated game which is supported by a self-consistent set of beliefs and actions, and shown that the grand coalition is an equilibrium coalition structure. However, this equilibrium is not unique.

3.4 An alternative equilibrium

The repeated game has one other symmetric equilibrium, which is induced by alternative beliefs. In this equilibrium, each player believes that all those players who announce C in Stage 1 will indeed form a coalition and choose their actions in Stage 2 so as to maximize their joint payoff, which will result in payoffs such as in (10); if two or more players announce NC in Stage 1, then the actions and payoffs will be the same as in the Nash equilibrium and the game will be repeated next period.

Given these alternative beliefs, let w again denote the equilibrium payoff or value of the game to each player. Let p_1,p_2, p_3 be the probabilities assigned by the three players to the C strategy. Then, in equilibrium, each player, say 3, should be indifferent between the strategies C and NC. Thus,

$$w = \tfrac{1}{3}p_1p_2 + \tfrac{1}{8}[p_1(1-p_2) + (1-p_1)p_2] + (1-p_1)(1-p_2)\beta w$$
$$w = \tfrac{3}{8}p_1p_2 + \beta w[p_1(1-p_2) + (1-p_1)p_2 + (1-p_1)(1-p_2)].$$

Using the symmetry requirement, i.e. $p_1 = p_2 = p_3 \equiv p$, we can solve for the equilibrium p and w from these two simultaneous equations. In fact, the two equations are equivalent to

$$\beta w = \tfrac{1}{8} - \tfrac{7}{28}p = \frac{\frac{1}{12}p^2 + \frac{1}{4}p}{\frac{1}{\beta} - (1-p)^2}.$$

It is easily seen that there is a unique solution $p > 0$ which is decreasing with β, and $p \to 0$ as $\beta \to 1$ implying that the probability of cooperation is very low if the discount factor β is close to 1. In this solution the value w to each player is no higher than $1/8$, which is smaller than the payoff of $1/3$ under the efficient equilibrium, and $w \to 1/8$ as $\beta \to 1$ and thus $p \to 0$.

Given this solution of the reduced Stage 1 game, we verify the ex-post optimality of the players' actions in Stage 2. Suppose all players happen to have chosen C, then it is clearly optimal for the players to give effect to the coalition and maximize their joint payoff as in (9). If two players, say 1 and 2, had chosen C and player 1 had chosen NC, then players 1 and 2 can get βw by deviating, i.e. by not forming the coalition, and thereby precipitating another round of play, but $1/8$ if they stick to the strategy and maximize their joint payoff as in (10). Since, as shown, βw is not higher than $1/8$, the deviation is not profitable. We have thus found another equilibrium of the repeated game which is also symmetric, but inefficient.

3.5 The γ-core imputation as a focal-point equilibrium

When a game has multiple equilibria, we must consider whether the players might be able to coordinate on a particular equilibrium by using information that is abstracted away from the game. Schelling (1960) argued that if the players' expectations can converge on one of the equilibria, then they may all expect it and hence implement it. According to Schelling, in a game with multiple equilibria, an equilibrium is a *focal-point* if it has some property that conspicuously distinguishes it from all other equilibria, and such an equilibrium is more likely to be observed. As noted earlier, the efficient equilibrium indeed has one such property, namely that it Pareto dominates the inefficient equilibrium, as the payoff of each player under the efficient equilibrium is strictly higher than under the inefficient equilibrium. 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. 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. We conclude that, given the restriction to symmetric equilibria and Schelling's focal-point argument, the formation of the grand coalition is the *only* equilibrium outcome of the repeated game.

4 Coalition formation in the general case

We now extend the analysis in the preceding section to the general model. Corollary 3 to Theorem 1 plays a crucial role in this extension. We assume that each player can choose more generally from among the set of 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 $(S_1, S_2, \dots, S_m)$, and a coalitional equilibrium as defined in (5).

As in the three-agent example, 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 to each player is $w = u_i^*$, as defined in (4). Thus, we only need to check whether the all-or-none strategy is ex post optimal. This means we must show that the players cannot gain by forming coalitions other than the grand coalition. We need the following lemma.

Lemma 4 *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 as members of 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 4 Let $(\tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_n)$ be the coalitional equilibrium corresponding to the coalition structure $P = (S_1, S_2, \dots, S_m)$. Then the first order conditions imply

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

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, since the players are identical, we must have $\tilde{u}_i < u_i^*$ for all $i \in S_k$ such that $|S_k| \geq |S_j|$ for all j.

Suppose now that the players try to test the all-or-none strategy by forming coalitions $(S_1, S_2, \ldots, S_m) \neq N$. Assume without loss of generality that $|S_1| \leq |S_2| \leq \cdots \leq |S_m|$. Consider the finite sequence of coalition structures $P^m \equiv (S_1, S_2, \ldots, S_m), P^{m-1} \equiv (S_1, S_2, \ldots, S_{m-1}, 1), P^{m-2} \equiv (S_1, S_2, \ldots, S_{m-2}, 1), \ldots, P^1 \equiv (S_1, 1)$. This sequence of coalition structures is obtained if the largest coalition in each subsequent coalition structure breaks up into singletons, starting from the largest coalition S_m in P^m. Let $\tilde{u}_i^k, i \in S_k, k = 1, 2, \ldots, m$ 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 4, $\tilde{u}_i^k < u_i^*$ for all $i \in S_k, k = 1, 2, \ldots, m$. Therefore,

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

for β sufficiently close to 1.

We begin by proving the assertion for the partition $P^1 = (S_1, 1), S_1 \neq N$ and then use induction to extend it to the partition $P = (S_1, S_2, \ldots, S_m)$. 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 (i.e. form singletons) 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 ex post optimal for the members of S_1 to break apart into singletons.

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. Then, $|S_2| \geq |S_1|$ and the payoffs of members of S_2 are $\tilde{u}_i^2, i \in S_2$. But if the members of S_2 follow the all-or-none strategy and break apart into singletons leading to the coalition structure $(S_1, 1)$, then, as just shown, the members of S_1 will also break apart into singletons and the game will be repeated, from which the expected payoff of each member i of coalition S_2 is $\beta u_i^* > \tilde{u}_i^2$, as in (12). Hence, the players in S_2 will break apart into singletons, and so will the players in S_1, as $\beta u_i^* > \tilde{u}_i^1, i \in S_1$.

By induction, it follows that, given the all-or-none expectation, the coalitions in the coalition structure $(S_1, S_2, \ldots, S_m)$ will break apart into singletons one-by-one, starting from the largest coalition S_m, until the finest coalition structure P^0 is reached.

This proves that the all-or-none strategy is ex post optimal and formation of the grand coalition is an equilibrium outcome of the repeated game. It is straightforward to show, as in the three agent example, that this equilibrium strictly Pareto dominates the inefficient mixed-strategy symmetric equilibrium. This comes from the fact that in the mixed-strategy symmetric equilibrium the probability for the grand coalition to form is strictly less than one, and therefore the payoff to each player is lower than in the symmetric efficient equilibrium. By the focal point argument, it follows that the grand coalition is the only equilibrium outcome of the repeated game.

4.1 The incentive to disintegrate

Disintegration of a coalition imposes negative externalities on other coalitions and weakens their incentives to form coalitions other than the grand coalition. As pointed out by Chander (1999), this can provide members of a farsighted coalition, incentives to disintegrate so as to induce the formation of the grand coalition later. We confirm this intuition in the following.

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 player i's payoff.

Proposition 5 *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 5 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 $i \in S_1$.

Examples are easily constructed in which $u_i(P) > u_i^* > u_i(P')$ for all $i \in S_1$ illustrating that disintegration of coalition S_2 may reverse the incentives of members of coalition S_1 to form a coalition other than the grand coalition. This is indeed true in the three-agent example when $S_2 = \{1, 2\}$ and $S_1 = \{3\}$.

If the coalition formation process reaches the finest coalition structure, then, as shown, it will result in the formation of the grand coalition. This provides members of farsighted coalitions who have lower payoffs compared to the payoffs they will get as members of the grand coalition an incentive to disintegrate as they know that this will ultimately lead to the finest coalition structure and therefore to the formation of the grand coalition.

4.2 Stability of the grand coalition

We first 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.

Theorem 6 *The grand coalition N is stable.*

Proof of Theorem 6 Suppose a coalition S breaks away from N. There are two possible cases: either $|S| \leq |N \backslash S|$ or $|S| > |N \backslash S|$. Let $P = (S, N \backslash S)$ and $P' = (S, 1)$.

Consider first $|S| \leq |N\backslash S|$. Then, in view of Lemma 4, $u_i(P) < u_i^*$ for all $i \in N\backslash S$ and, by Corollary 3 to Theorem 1, $u_i(P') < u_i^*$ for all $i \in S$. This means that, given the coalition structure $P = (S, N\backslash S)$, the members of $N\backslash S$ stand to gain if they break apart, since that would then induce the members of S also to break apart and lead to the formation of the grand coalition. This proves that coalition S cannot strictly improve its payoff by breaking away from N.

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

This leaves out the possibility that rather than breaking up into singletons, $N\backslash S$ or S may break up into two or more non-singleton coalitions. However, there is no loss of generality in ignoring this possibility, since as shown that would only lead to intermediate coalition structures in which the coalitions will continue to break apart until the finest coalition structure is reached and the grand coalition N is formed.

5 Conclusions

The motivation for the analysis in this paper comes from the observation that which coalitions will form from the finest coalition structure determines which coalition structures are stable. This leads to a formulation of the problem of coalition formation from the finest coalition structure as an infinitely repeated game. It is shown that a certain equilibrium of this game corresponds to the γ-core assumption that when a coalition forms the remaining players form singletons, and that the grand coalition is an equilibrium coalition structure. Both these results are quite robust with respect to the assumptions. First, as noted earlier, the requirement that the equilibrium be symmetric is not necessary. Second, it is unnecessary to allow mixed strategies, as the game has a dominant strategy equilibrium. Third, the agents need not be identical. For reasons of space, we do not present this generalization here, but the argument for it is as follows.

The main difficulty in generalizing the results to the case of asymmetric agents is that Lemma 4 does not hold when the players are not identical, since their payoffs in that case neither depend on the size of the coalitions nor are comparable across coalitions. But it can be shown, under a rather mild condition, that given an imputation of the γ-core and any coalition structure $P = (S_1, S_2, \ldots, S_m), P \neq P^0, N$, there exists a non-singleton coalition S_i (which is not necessarily the largest) in P such that at least one of its member's payoff is lower than his payoff in the grand coalition. If such a player leaves the coalition S_i then the new coalition structure will similarly have a non-singleton coalition and a member of that coalition whose payoff is lower and so on. All non-singleton coalitions will therefore continue to disintegrate until the finest coalition structure is reached if they believe that that would lead to the formation of the grand coalition later. This result, in lieu of Lemma 4, is sufficient to establish the ex post optimality of the all-or-none strategy and that the grand coalition is an equilibrium coalition structure.

However, the stronger result that the grand coalition is the *only* equilibrium coalition structure is not similarly robust. The restriction to symmetric equilibria is critical.[15] If the equilibria are not required to be symmetric, then, as shown by Ray and Vohra (1997) and Yi (1997), beside the grand coalition, some inefficient asymmetric coalition structures also will obtain in equilibrium.

Another assumption that limits our analysis is that the players choose their strategies simultaneously. Bloch (1996) and Ray and Vohra (1999, 2001) consider sequential games of coalition formation in which the players choose their strategies according to a pre-specified order of play. This gives some players the ‘first mover advantage’ and therefore the argument for restricting to symmetric equilibria no longer applies. Accordingly, Bloch (1996) and Ray and Vohra (1999, 2001) obtain inefficient asymmetric coalition structures as the equilibrium outcomes of the sequential game.

Finally, we have restricted our analysis to a class of games that imply positive externalities from coalition formation. For games with negative externalities, the analogous assumption underlying the γ-characteristic function is that the deviating coalition S assumes that the residual coalition $N\backslash S$ will form and its members will adopt the best response *joint* strategies. However, we do not pursue the analysis of this case here as it requires a new model in which coalition formation imposes negative externalities.

Acknowledgements I thank Anindya Bhattacharya, Indranil Chakraborty, Henry Tulkens, and two anonymous referees for helpful comments and CORE for its hospitality.

References

Barrett S (1994) Self-enforcing international environmental agreements. Oxf Econ Pap 46:804–878

Bloch F (1996) Sequential formation of coalitions in games with externalities and fixed payoff division. Games Econ Behav 14:90–123

Carraro C, Siniscalco D (1993) Strategies for the international protection of the environment. J Pub Econ 52:309–328

Chander P (1993) Dynamic procedures and incentives in public good economies. Econometrica, 61:1341–1354

Chander P, Tulkens H (1995) A core-theoretic solution for the design of cooperative agreements on transfrontier pollution. Int Tax Pub Financ 2:279–293

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

Chander P (1999) International treaties on global pollution: a dynamic time-path analysis. In: Ranis G, Raut LK (eds) Trade, growth and development: essays in honor of T.N. Srinivasan. Elsevier, Amsterdam

D’Aspremont C, Gabszewicz JJ (1986) On the stability of collusion. In: Stiglitz JE, Mathewson GF (eds) New developments in the analysis of market structure. MIT Cambridge, pp. 243–264

Diamantoudi E, Xue L (2002) Coalitions, agreements and efficiency. mimeo, University of Aarhus

[15] The justification for this restriction is that the symmetric equilibria are the only ones that give each identical player an equal chance to be a free rider.

Foley D (1970) Lindahl solution and the core of an economy with public goods. Econometrica 38:66–72
Helm C (2001) On the existence of a cooperative solution for a coalitional game with externalities. Int J Game Theory 30:141–147
Holmstrom B, Nalebuff B (1992) To the raider goes the surplus? A re-examination of the free-rider problem. J Econ Manage Strat 1:37–62
Mäler K-G (1989) The acid rain game. In: Folmer H, van Ierland E (eds) Valuation methods and policy making in environmental economics. Elsevier, Amsterdam, pp. 231–252
Mauleon A, Vannetelbosch V (2004) Farsightedness and cautiousness in coalition formationgames with positive spillovers. Theory Dec Sci 56:291–324
Moulin H (1987) Egalitarian-equivalent cost sharing of a public good. Econometrica 55:963–976
Ray D, Vohra R (1997) Equilibrium binding agreements. J Econ Theory 73:30–78
Ray D, Vohra R (1999) A theory of endogenous coalition structures. Games Econ Behav 26:286–336
Ray D, Vohra R (2001) Coalitional power and public goods. J Polit Econ 109:1355–1384
Roberts DJ (1974) The Lindahl solution for economies with public goods. J Pub Econ 3:23–42
Scarf H (1971) On the existence of a cooperative solution for a general class of person games. J Econ Theory 3:169–181
Schelling TC (1960) The strategy of conflict. Harvard University Press, Cambridge
Starrett D (1972) A note on externalities and the core. Econometrica 41:179–183
Thrall R, Lucas W (1963) N-person games in partition function form. Naval Res Logist Qu 10:281–198
Tulkens H (1998) Cooperation versus free-riding in international environmental affairs: two approaches. In: Hanley N, Folmer H (eds) Game theory Environ Edward Elgar, UK, pp. 30–44
Yi S (1997) Stable coalition structures with externalities. Games Econ Behav 20:201–237