DISCUSSION PAPER SERIES

No. 3258

STABLE COALITIONS

Carlo Carraro and Carmen Marchiori

INTERNATIONAL TRADE

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP3258.asp**

ISSN 0265-8003

STABLE COALITIONS

Carlo Carraro, Università di Venezia and CEPR
Carmen Marchiori, Fondazione ENI Enrico Mattei

Discussion Paper No. 3258
March 2002

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 3258

March 2002

ABSTRACT

Stable Coalitions*

This Paper examines recent theoretical developments in the theory of coalition stability. It focuses on the relationship between the incentives to defect from a coalition, the size of the resulting equilibrium coalition structure and the different assumptions on membership rules, coalition behaviour, players' conjectures, etc. The Paper considers several cases. Simultaneous versus sequential moves, linear versus circular order of moves, Nash versus rational conjectures, open versus exclusive membership, monotonic versus non-monotonic pay-off functions and orthogonal versus non-orthogonal reaction functions. The profitable and stable coalition will be derived for each possible configuration of the rules of the game, the pay-off functions and the membership rules. The results show that the size of the profitable and stable coalition highly depends on the chosen configuration and that the equilibrium outcome ranges from a small coalition with a few signatories to full cooperation. The Paper explores under which conditions a large stable coalition is likely to emerge, and identifies the institutional setting that favours the emergence of such coalition.

JEL Classification: H00, H20 and H30
Keywords: agreements, coalitions, incentives, negotiations and stability

Carlo Carraro
Dipartimento di Scienze Economiche
Università di Venezia
San Giobbe 873
30121 Venezia
ITALY
Tel: (39 041) 257 4166
Fax: (39 041) 257 4176
Email: ccarraro@unive.it

Carmen Marchiori
FEEM
Università di Venezia
San Giobbe 873
30121 Venezia
ITALY
Email: cmarchiori@unive.it

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=104121

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=157262

* A previous version of this Paper, titled ‘Endogenous Formation of Environmental Coalitions’ and co-authored with F Moriconi, was presented at the Coalition Theory Network Workshop on ‘Coalition Formation: Applications to Economic Issues’, Venice, 8–10 January 1998, at the University of Paris X Seminar in 1999 and at the 2000 EAERE-FEEM-VIU Summer School. The authors are grateful to Sergio Currarini, Gilles Rotillon, Tarik Tazdait, the CTN workshop participants and the students of the Summer School for helpful comments and suggestions.

Submitted 09 February 2002

STABLE COALITIONS

1. Introduction

Coalitions are a widespread phenomenon in old and modern societies. Households can be seen as coalitions, as well as nations or research joint ventures. In industrial economics, cartels have been interpreted as coalitions, whereas in environmental economics, agreements among several nations to protect the international environment have also been described as coalitions.

In recent years, the problem of global public goods, such as the international environment, but also peace or crime control, has become increasingly important. Trans-national externalities, being related to knowledge, information, migrations, pollution, crime and similar flows, play a major role in international politics. Their importance and close link to economic development may affect growth and welfare at a macroeconomic level. Hence, the need for global policies to correct their undesired effects. In the present institutional setting - with no supranational governing bodies - global externalities cannot be corrected by global institutions, which do not exist, but need to be corrected by voluntary agreements among sovereign states and/or by voluntary initiatives undertaken by firms, industries, households, NGOs, etc. This situation requires international cooperation among nations and/or agents.

To analyse the emergence of international cooperation, economists have used a fairly new game-theoretic approach, in which cooperation is the outcome of a non-cooperative strategic behaviour of the players involved in the negotiations. This approach has been proposed both in games without spillovers (Le Breton and Weber, 1993; Konishi, Le Breton and Weber, 1997) and in games with positive or negative spillovers (Bloch, 1996, 1997; Carraro and Siniscalco, 1993; Yi, 1997).[1]

The game is therefore a two-stage game: in the first stage, countries/agents decide non cooperatively whether or not to sign the agreement (join the coalition) given the adopted burden-sharing rule; in the second stage, countries/agents set their policy/decision variables by maximising their welfare function given the decision taken in the first stage and the adopted burden-sharing rule. Signatories, i.e. those countries/agents which decide non cooperatively to sign the agreement/join the coalition, act cooperatively (i.e. they maximise a joint welfare function), whereas non signatories play a non-cooperative Nash game against the group of

[1] It must be acknowledged that the first results on the non-cooperative formation of coalitions in the presence of positive spillovers can be found in the oligopoly literature on stable cartels. See D'Aspremont and Gabszewicz (1986), Donsimoni, Economides and Polemarchakis (1986).

signatories. These behavioural assumptions are quite standard and formalised in the concept of game γ-equilibrium by Chander and Tulkens (1995, 1997).[2]

This paper aims at analysing the above game from different viewpoints. The ultimate goal is to determine the stable coalition structures of the game and whether a non-trivial coalition emerges at the equilibrium. This can be the grand-coalition or, more frequently, a partial coalition whose members are a subset of all players of the game. More importantly, we would also like to examine how the stable coalition structures change when assumptions on the rules of the game change. In particular, we will account for simultaneous vs. sequential games, linear vs. circular orders of moves, open vs. exclusive vs. coalition unanimity membership rules, orthogonal vs. non-orthogonal free-riding, monotonic vs. humped-shaped payoff functions, Nash vs. rational conjectures. In all cases, the equilibrium concept used to analyse the emergence of cooperation is strictly non-cooperative. Moreover, we focus on coalition structures in which only one non-trivial coalition belongs to the equilibrium. This case is the one most frequently considered in the economic literature, because in most situations a single agreement is proposed and countries/agents decide whether or not to sign it.[3]

As a consequence, one of the outcomes of the paper is a sort of taxonomy of equilibria of the non-cooperative game in which countries decide whether or not to join a coalition. This taxonomy will hopefully include all cases already discussed in the economic literature and also provide a consistent analysis of new types of coalition games. Note that the economic situation which provides a background and a justification to our analysis is mostly the provision of global public goods and the management of global commons. However, our results can be applied to several other situations, both in economics and political science.

The structure of the paper is as follows. Section 2 provides the main definitions and assumptions. Section 3 defines the concepts of spillovers and free-riding which will be considered in the subsequent analysis and introduces our definitions of profitability and free-riding functions. Sections 4 and 5 identify the stable coalition structures and how stable coalitions change according to the rules of the game. Some concluding remarks are provided in section 6.

---

[2] A notable exception is the paper by Barrett (1994) where the coalition is assumed to be the leader of a Stackelberg game.

[3] The possibility of equilibria with multiple coalitions is analysed in Carraro (1998, 1999, 2000) and in Finus and Rundhagen (2001). This possibility is also considered in Bloch (1997), Yi (1997), Ray and Vohra (1999).

2. Assumptions and Definitions

Assume negotiations take place among n countries/agents, $n \geq 3$, each indexed by i = 1, ... , n. Countries/agents play a two-stage game. In the first stage -- the *coalition game* -- they decide non-cooperatively whether or not to sign the agreement (i.e. to join the coalition *c* of cooperating countries). In the second stage, they play a non cooperative *policy game* in which countries/agents who have signed the agreement play as a single player against non signatories. Signatories divide the resulting payoff according to a given burden-sharing rule (any of the rules derived from cooperative game theory). As said, these behavioural rules coincide with those underlying the concept of γ-equilibrium in Chander and Tulkens (1995, 1997).

This two stage game can be represented as a *game in normal form* denoted by $\Gamma = (N, \{X_i\}_{i \in N}, \{u_i\}_{i \in N})$, where N is a finite set of players, X_i the strategy set of player i and u_i the payoff function of player i, assigning to each profile of strategies a real number, i.e. $u_i : \Pi_{i \in N} X_i \rightarrow R$. The payoff function is a twice continuously differentiable function.

A *coalition* is any non-empty subset of N. A *coalition structure* $\pi = \{C_1, C_2, \ldots, C_m\}$ is a partition of the player set N, i.e. $C_i \cap C_j = \varnothing$ for $i \neq j$ and $\bigcup_{i=1}^{m} C_i = N$. In international games, as well as in many other games, the formation of coalitions creates externalities. Hence, the appropriate framework to deal with this situation is a *game in partition function form*, in which the payoff of each player depends on the entire coalition structure to which he/she belongs (Bloch, 1997; Ray and Vohra, 1999). This is why we convert the game in normal form into a game in partition function.

Denote by Π the set of all feasible coalition structures. A *partition function* P: $\Pi \rightarrow R$ is a mapping which associates to each coalition structure π a vector in $\Re^{|\pi|}$, representing the worth of all coalitions in π. In particular, $P(C_i; \pi)$ assigns to each coalition C_i in a coalition structure π a worth. When the rule of payoff division among coalition members is fixed, the description of gains from cooperation is done by a *per-member partition function* p: $\Pi \rightarrow \Re^n$, a mapping which associates to each coalition structure π a vector of individual payoffs in $\Re^n$. In particular $p(C_i; \pi)$ represents the payoff of a player belonging to the coalition C_i in the coalition structure π.

Under suitable assumptions (A.1 and A.2 below), the second stage of the game can be reduced to a partition function (Yi, 1997; Bloch 1997). Therefore, the study of coalition stability consists of the study of the first stage of the game, i.e. the one in which players decide whether or not to sign the agreement. A necessary assumption for the second stage of the game to be reduced to a partition function is the following one.

A.1 (Uniqueness). The second stage game, in which all players decide simultaneously, has a unique Nash equilibrium for any coalition structure.

Moreover, in order to convert the strategic form into a partition function one, the competition among the various coalitions must be specified. As said, the most common, and perhaps the most natural assumption, is:

A.2 (g game). Inside each coalition players act cooperatively in order to maximise the coalitional surplus, whereas coalitions (and singletons) compete with one another in a non cooperative way.

The partition function is then obtained as a Nash equilibrium payoff of the game played by coalitions and singletons. Formally, for a fixed coalition structure $\pi = \{C_1, C_2, \dots, C_m\}$, let x* be a vector of strategies such that:

$$\forall C_i \in \pi, \qquad \sum_{j \in C_i} u_j\left(x^*_{C_i}, x^*_{N \setminus C_i}\right) \geq \sum_{j \in C_i} u_j\left(x_{C_i}, x^*_{N \setminus C_i}\right) \qquad \forall x_{C_i} \in \times_{j \in C_i} X_j$$

Then define:

$$P(C_i; \pi) = \sum_{j \in C_i} u_j\left(x^*\right).$$

In order to simplify the derivation of the partition function, we introduce a further assumption:

A.3 (Symmetry). All players are ex-ante identical, which means that each player has the same strategy space in the second stage game.

This assumption allows us to adopt an equal sharing payoff division rule inside any coalition, i.e. each player in a given coalition receives the same payoff as the other members.[4] Furthermore, the symmetry assumption implies that a coalition C_i can be identified with its size c_i and a coalition structure can be denoted by $\pi = \{c_1, c_2, \dots, c_m\}$, where $\sum_i c_i = n$. As a consequence, the players payoff depends only on the coalition sizes and not on the identities of the coalition members. The per-member partition function (partition function hereon) can thus be denoted by $p(c_i; \pi)$, which represents the payoff of a player belonging to the size-c_i coalition in the coalition structure π.

We also assume:

---

[4] We consider the equal sharing rule as an assumption, because it is not endogenously determined in the model. However Ray and Vohra (1999) provides a vindication for this assumption.

A.4 (Single coalition). Coalition structures are characterised by a single coalition. Non-cooperating players behave as singletons.

This assumption, which is adopted in most analyses of international games, allows us to further simplify previous definitions. A coalition structure can indeed be denoted by $\pi = \{c, 1_{(n-c)}\}$ and the per-member partition function becomes $p(c; \pi)$.

Let us now identify the main features of the game. The decision process through which a coalition forms can be modelled either as a *simultaneous* game, in which all players announce at the same time their strategic choice, or as a *sequential* one, in which each player can announce his/her strategy according to an exogenous order of moves. In this paper, we will consider both a *linear* and a *circular* order of moves. In the three cases – simultaneous, linear sequential, circular sequential -- the ruling solution concept is (almost always) some variant of the non-cooperative Nash equilibrium, because we focus mainly on coalition stability in games where self-interested players can only agree on strategy profiles consistent with the best-reply property.

Our classification of coalition games also includes three different membership rules: the open membership rule, the exclusive membership rule and the coalition unanimity rule. In the *open membership game*[5] (D'Aspremont *et al*. 1983; Carraro and Siniscalco, 1993; Yi and Shin, 1994) each player is free to join and to leave the coalition without the consensus of the other coalition members. The decision process can be thought of as if each player announces a message. In this game, all players who announce the same message form a coalition. This open membership rule thus implies that a coalition accepts any new player who wants to join it.

By contrast, in the *exclusive membership game*[6] (Yi and Shin, 1994) or *game* Δ[7] (Hart and Kurz, 1983), each player can join a coalition only with the consensus of the existing members, but he/she is free to leave the coalition. In this decision process, each player's message consists in a list of players with whom he/she wants to form a coalition. Then, players who announce the same list form a coalition, which is not necessarily formed by all players in the list.

---

[5] The open membership game is "a game in which membership in a coalition is open to all players who are willing to abide by the rules of the coalition" (Yi and Shin, 2000).

[6] The exclusive membership game is a "a game in which the existing members in the coalition are allowed to deny membership to outsider players" (Yi and Shin, 2000).

[7] The game Δ is "a game in which the choice of a strategy by a player means the largest set of players he is willing to be associated with in the same coalition. Each set of all the players who chose the same C then forms a coalition (which may, in general , differ from C)" where C indicates a subset of players (Hart and Kurz, 1983).

Finally, in the *coalition unanimity game*[8] (Chander and Tulkens, 1993; Yi and Shin, 1994; Bloch, 1997) or *game G* (Hart and Kurz, 1983)[9], no coalition can form without the unanimous consensus of its members. This implies that players are not free either to join the coalition or to leave it. In this decision process, players' messages consist in a list of players as in the previous one. However, if a coalition is formed, it is necessarily composed of all players in the list and as soon as a player defects the coalition breaks up into singletons.

Another feature of the game with important implications on the equilibrium coalition structure is the form of per-member partition function. It is also useful to simplify the distinction between the membership rules described above. Two cases will be considered. In the first one, the per-member partition function is positively sloped and *monotonic*, whereas in the second one it is *humped-shaped*. In the monotonic case, and above a minimum coalition size c^m (see below), the per-member partition function increases monotonically with respect to the number of signatories of the agreement. Formally, $p(c; \pi)$, where $\pi = \{c, 1_{(n-c)}\}$, is an increasing function of c for all $c \geq c^m$, $c^m \geq 2$. In the humped-shaped case, and above a minimum coalition size c^m, there is an optimal size $c°<n$ at which the per-member partition function is maximised. Formally, $p(c; \pi)$, where $\pi = \{c, 1_{(n-c)}\}$, increases with c in the interval $[c^m,c°)$ and decreases with c for $c° < c \leq n$.

As a consequence, in the monotonic case, the open membership and the exclusive membership rules are equivalent, because no players in the coalition have an incentive to ban the entry of other players. The analysis of coalition stability is thus limited to two rules (open membership and coalition unanimity). By contrast, when the per-member partition function is humped-shaped, we need to consider all three rules because, for $c > c°$, coalition members have an incentive to prevent others players from joining the coalition.

Both in the case of monotonic and humped-shaped per-member partition function, we make a final assumption which simplifies our analysis.

A.5 (Marginal gains). The marginal per member partition function $\partial p(c; \pi)/\partial c$ *is smaller than* $\partial p(1; \pi)/\partial c$ -- *the marginal payoff function of free-riders -- for all 2 £ c £ n.*

This assumption is based on the idea that even when members of a coalition increasingly gain from additional memberships, the marginal gain of free-riders is even larger. The intuition is as follows. Even when cooperating players gain from increasing the coalition size, they still pay a cost to obtain the coalition benefits. By contrast, free-riders obtain the same benefits (as said in the Introduction, we are focusing mainly

---

[8] The coalition unanimity game is a game in which "a coalition forms if and only if all potential members agree to join it" (Yi and Shin, 1994).

[9] The game Γ is a game in which "each player chooses the coalition to which he wants to belong. A coalition forms if and only if all its members have in fact chosen it; the rest of the players become singletons" (Hart and Kurz, 1983).

on the case of global public goods) without paying any costs. Assuming an increasing cost function, the marginal gain for coalition members provided by an additional player in the coalition is equal to the additional benefit minus the additional cost. In the case of free-riders, the marginal gain is larger, because it is equal to the additional benefit without any costs.

Note that Assumption 5 implies that, if $p(c; \pi)$ is monotonically increasing, then $p(1; \pi)$ is also monotonically increasing. If $p(c; \pi)$ is positive for all $c \in [2, n]$, then $p(1; \pi)$ is also positive for all $c \in [2, n]$.

Another important element of our taxonomy is the type of conjecture formed by players when taking their decision about joining or leaving a coalition. We consider two cases. The usual *Nash conjecture* for which a player *i* takes his decision given the other players' decisions, which are assumed not to change as a consequence of player *i*'s decision. This implies that when a player decides to defect from a coalition, he/she conjectures that the other players will keep cooperating. By contrast, in the *rational conjecture* case, a player leaving a coalition can rationally predict the consequences of his/her decision and therefore how many players will then decide to join or leave the coalition as a consequence of his/her decision to leave the coalition.

In our taxonomy, we do not discuss the case in which spillovers from one coalition to another coalition are positive and one in which they are negative.[10] We focus only on positive spillovers. As said above, many economic problems are characterised by the incentive to free-ride on the other players' cooperative behaviour. In other words, a coalition provides *positive spillovers* to the other players or coalitions. The case of positive spillovers is also one in which coalitions are most likely to be unstable. Hence, it provides a benchmark for all other possible assumptions on coalition spillovers.

Table 1 summarises the taxonomy that we adopt to survey the main contributions to the theory of coalition stability and to provide new results. For each combination of the three main features of the game – order of moves, membership rules, conjectures – we discuss the case in which the players' reaction functions are *orthogonal* vs. the case in which they are *negatively sloped*. The orthogonal case is particularly important in international environmental games, because it implies that free-riders enjoy the cleaner environment produced by the emission abatement carried out in some countries, but do not expand their own emissions. By contrast, the non-orthogonal case is the one in which market mechanisms induce the so-called "leakage" problem, i.e. free-riders benefit twice from the other countries' cooperation, both because they get a cleaner environment and because they can increase their degree of exploitation of natural resources.

---

[10] See Yi (1997) for this distinction.

Table 1. Coalition Games. Taxonomy

	Open membership (monotonic or humped-shaped payoff)	**Exclusive membership** (humped-shaped payoff)	**Coalition unanimity** (monotonic or humped-shaped payoff)
Simultaneous game (orthogonal or negatively sloped reaction funct.)	Nash or Rational conjectures	Nash or Rational conjectures	Nash
Linear sequential game (orthogonal or negatively sloped reaction funct.)	Subgame perfect	Subgame perfect	Subgame perfect
Circular sequential game (orthogonal or negatively sloped reaction funct.)	Subgame perfect	Subgame perfect	Subgame perfect

3. Single coalition games with positive spillovers

As assumed above, in this paper we consider the situation in which countries are proposed to sign a *single agreement*. From a game-theoretic viewpoint, this implies that only one coalition can be formed, the remaining defecting players playing as singletons. The first stage of the game is therefore a binary choice game (joining the coalition or behaving as a lone free-rider) and the outcome of this interaction is a single coalition structure $\pi = \{c,1_{(n-c)}\}$.

In this setting, we can provide a formal definition of positive spillovers and free-riding.

<u>Positive spillovers</u>. In any single coalition structure, if some players form a coalition, the other players are better off. Then the partition function of any player outside the coalition -- the non-member partition function -- is increasing in c for all values of c. Formally, $p(1; \pi)$, where $\pi = \{c, 1_{(n-c)}\}$, is an increasing function of c.

The existence of positive spillovers creates an incentive to free ride on the coalition decision to cooperate. In particular, two different free-riding behaviour patterns may emerge according to the slope of players' reaction functions (Carraro and Siniscalco, 1993).

Orthogonal free-riding. When players' reaction functions in the policy game are orthogonal, free-riders benefit from the cooperative abatement of the coalition, but have no incentive to damage it (in international environmental games, there is no "leakage").

Non orthogonal free-riding. When players' reaction functions in the policy game are non-orthogonal, i.e. in an environment in which there is a greater interdependence between countries' policy strategies, free-riders can damage the coalition (e.g. in environmental games, by increasing emissions whenever cooperating countries reduce their own). In this case, free-riding can lead to decreasing returns from cooperation (in particular for small coalitions). This implies that a small number of cooperators may loose from cooperation, because of the strategy adopted in the free-riding countries/agents. Let c^m be the size a coalition has to reach in order to start benefiting from cooperation and to obtain an increasing benefit as the coalition becomes larger and larger.[11]

As stated in the Introduction, this paper focuses on coalition games characterised by positive spillovers and incentives to free ride, a framework which is useful to analyse many international economic and political problems. Given these features of the n-player coalition game, let us characterise its equilibrium. As said, players decide independently and non-cooperatively whether or not to join the coalition. An obvious necessary condition for the existence of a coalition is its profitability.

Profitability. A coalition c is profitable if each cooperating player gets a payoff larger than the one he/she would get when no coalition is formed. Formally:

(1) $$p(c; \pi) > p(1; \pi^S)$$

for all players in the coalition c, where $\pi = \{c, 1_{(n-c)}\}$ and $\pi^S = \{1_n\}$.

Figure 1 shows the profitability function $P(c) = p(c; \pi) - p(1; \pi^S)$ for the two types of free-riding behaviour that we consider in this paper. Notice that, in the presence of non orthogonal free-riding, P(c) is positive for values of c above c^m. In a similar way, we can define the gain from free-riding on the coalition cooperative behaviour. If a coalition c forms, a free-rider achieves $Q(c) = p(1; \pi) - p(1; \pi^S)$, where Q(c) is the free-riding function. This is shown in Figure 2, where two cases are identified. The case in which free-riding is orthogonal and fixed abatement costs are low, and the case in which free-riding is strongly non-orthogonal, thus inducing decreasing returns from free-riding.[12]

---

[11] For example, in the case of climate change, both situations are possible. Current estimates of "carbon leakage" ranges from very small values to quite large ones (Ulph, 1993). Therefore, it is worth analysing both cases and the impact of different degrees of "carbon leakage" on the outcome of negotiations on climate change control.

[12] Botteon and Carraro (1997b) provide some numerical examples of this situation.

Figure 1. Member partition function

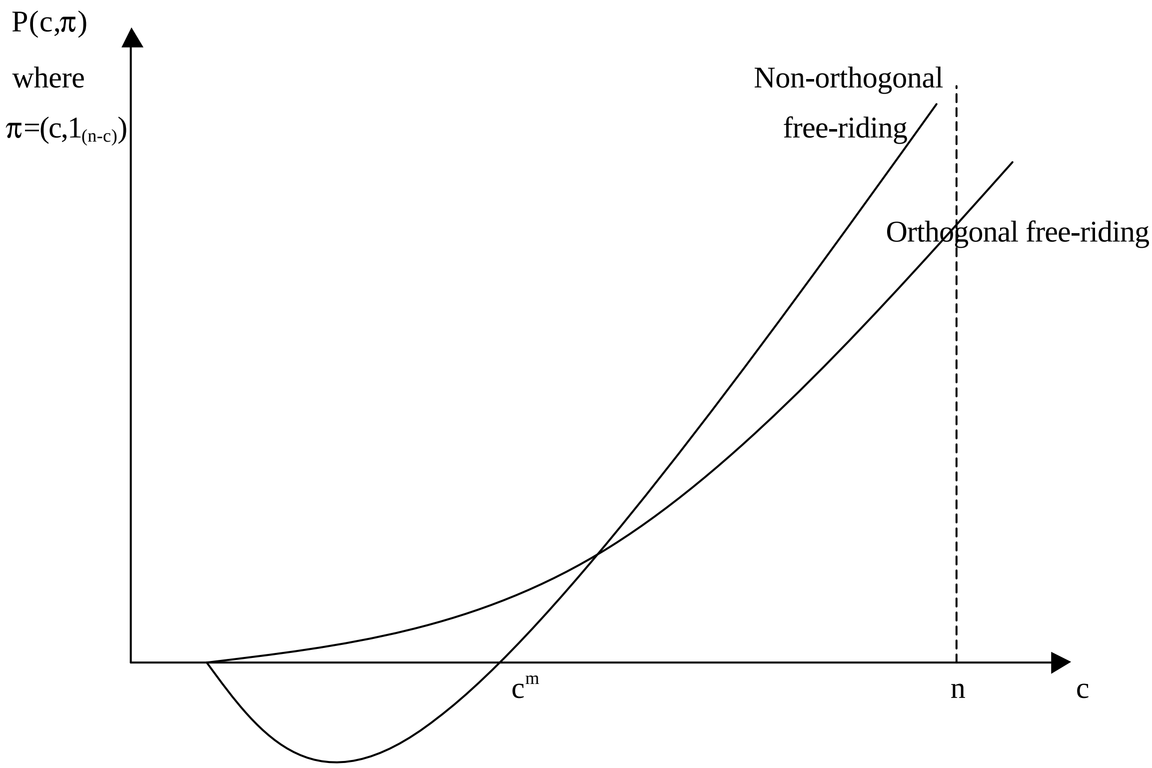
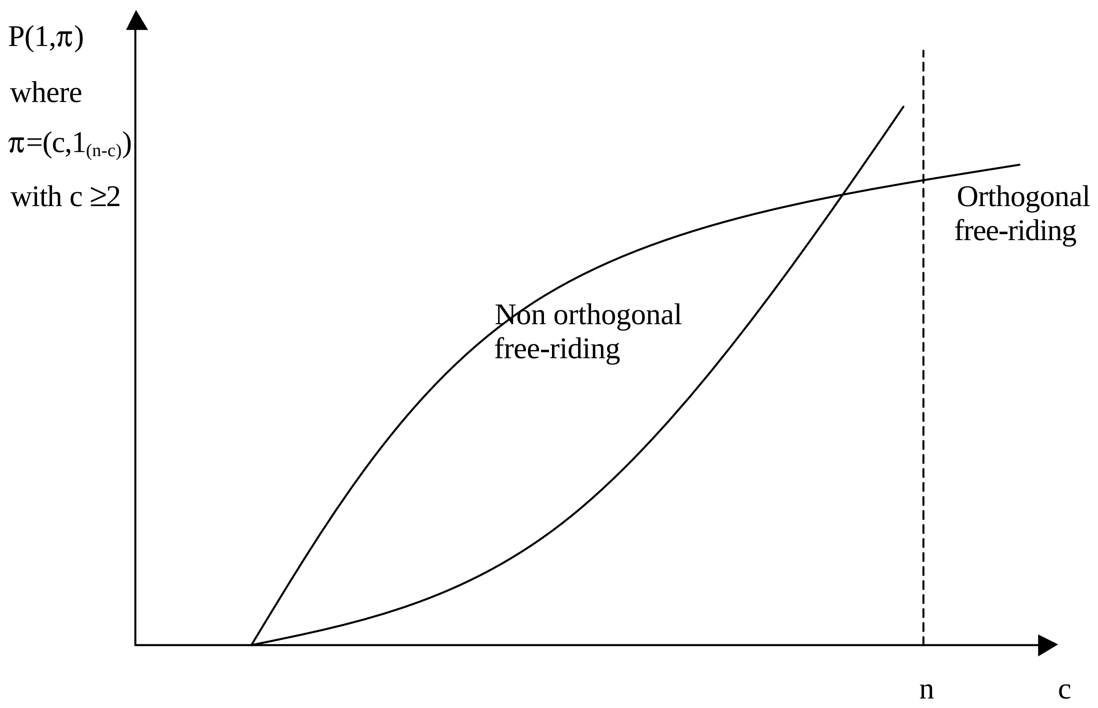


Figure 2. Non-member partition function



From (1), the value of the minimal profitable coalition size c^m can easily be derived. Indeed, c^m is the value of c at which P(c) becomes positive. This value depends on the strategic interaction between the coalition and the singleton players. In particular, with orthogonal free-riding, any coalition size is profitable and c^m is simply 2, i.e. P(c) is non-negative for all $c \geq 2$ (Carraro and Siniscalco, 1992). By contrast, with non-orthogonal free-riding, the coalition has to reach a minimal size by which it can offset the damaging action undertaken by free-riders. This size, denoted by c^m, is generally larger than two (see Figure 1). In the sequel, we assume $c^m \leq n$, i.e. the grand coalition n is profitable.

The profitability condition defined above is only a necessary condition for an equilibrium coalition to form. In order to completely characterise the equilibrium of the game, we need to introduce the concept of stability. This concept differs with the rules of the game. Therefore, in the following sections, we will introduce a definition of stability for each set of rules. Let us start the analysis with simultaneous coalition games.

4. Simultaneous games

Let us assume that in the first stage of the game all players decide simultaneously whether or not to join the coalition. Let us then analyse the equilibria of the coalition game for three different membership rules. In all propositions below, Assumptions 1-5 will be used.

4.1 Open membership

This is the membership originally adopted in the cartel stability literature (D'Aspremont et al, 1983; D'Aspremont and Gabsewicz, 1986) and in the environmental literature on international agreements (Hoel, 1991; Carraro and Siniscalco, 1992, 1993; Barrett, 1994). Under this rule, there is no restriction to entry and exit of players from the coalition. Moreover, let us assume that:

4.1.1 Nash conjectures

In the first stage of the game, each player outside the coalition takes as given the decision of the other players or coalition; each player inside the coalition maximises the joint welfare taking as given the decision of the free-riders. The equilibrium of the coalition game is thus a Nash equilibrium, completely characterised by the following condition:

Nash Stability (*D'Aspremont et al, 1983*). A coalition is stable if it is both internally and externally stable. It is internally stable if no cooperating player is better off by defecting in order to form a singleton. Formally:

(2a) $p(c; \pi) \geq p(1; \pi')$,

where $\pi = \{c, 1_{(n-c)}\}$ and $\pi' = \pi \backslash \{c\} \cup \{c-1, 1\}$ for all $i \in c$.[13] It is externally stable if no singleton is better off by joining the coalition c. Formally:

(2b) $\quad p(1; \pi) > p(c'; \pi'),$

where $\pi = \{c, 1_{(n-c)}\}$ and $\pi' = \pi \backslash \{c, 1\} \cup \{c'\}$ and $c' = c+1$ for all $i \notin c$.

The stability function (*Carraro and Siniscalco, 1992*). The stability function is a useful tool to identify the size of a (Nash) stable coalition, i.e. the equilibrium of the simultaneous coalition game with open membership. Let:

(3) $\quad L(c) = p(c;\pi) - p(1;\pi') = P(c) - Q(c-1)$

be the stability function, where $\pi = \{c, 1_{n-c}\}$, $\pi' = \{c', 1_{n-c'}\}$ and $c' = c-1$.

When positive, the stability function shows that a singleton has an incentive to join the coalition c. When negative, it signals an incentive to free-ride on the coalition action. The stable coalition size is the one where no cooperating player is willing to defect and no free-rider is willing to join the coalition. Hence, as formally shown below, it coincides with the largest integer c* below the value of c for which $L(c) = 0$ and $L'(c) \leq 0$.

A few diagrams will help us to identify possible equilibrium coalition sizes. By definition, the function Q(c-1) is equal to the non-cooperative payoff when c=2, whereas the member partition function P(c) is equal to the non-cooperative payoff when c=1. Figure 3, 4, 5 and 6 show the functions P(c) and Q(c-1) and the resulting function L(c) for different marginal incentives to cooperate and free-ride (using Assumption 5). These figures capture four main situations that are likely to occur, but are certainly not exhaustive of all possible cases. Their analysis is however helpful to understand which mechanisms support a stable coalition.

Stability function with orthogonal free-riding. Consider first Figure 3, i.e. the case in which the profitability function is monotonic. This case has been studied in several environmental economics contributions (Cf. Carraro and Siniscalco, 1992, Chander and Tulkens, 1997) and in simple models of public goods provision (Yi, 1997). In this case, the game is characterised by positive spillovers. There is no feedback because there is no leakage. As a consequence, both the per-member function and the non-member one are monotonically increasing for all values of c. The two curves – P(c) and Q(c-1) -- intersect at c* < n, i.e. the stability

[13] We assume that if a player is indifferent between joining the coalition or free-riding, then he/she joins the coalition.

function L(c) becomes negative for coalition sizes above c*. The coalition size c* defines the Nash-stable coalition (see Proposition 1). If c* is larger than one, the stable coalition structure is non-trivial.

Figure 3. Stability function with orthogonal free-riding and monotonic payoff

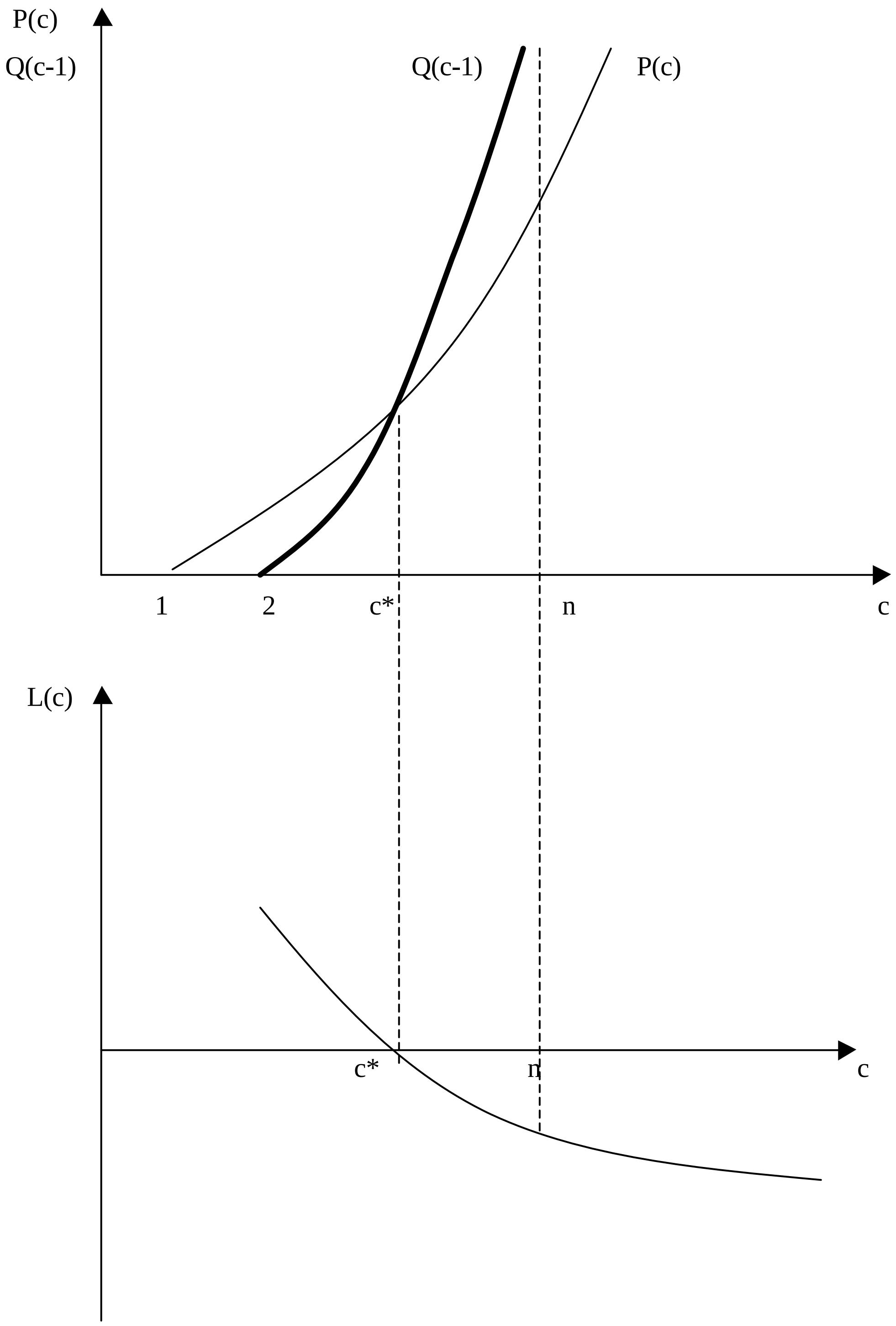


Figure 4. Stability function with orthogonal free-riding and humped shaped payoff

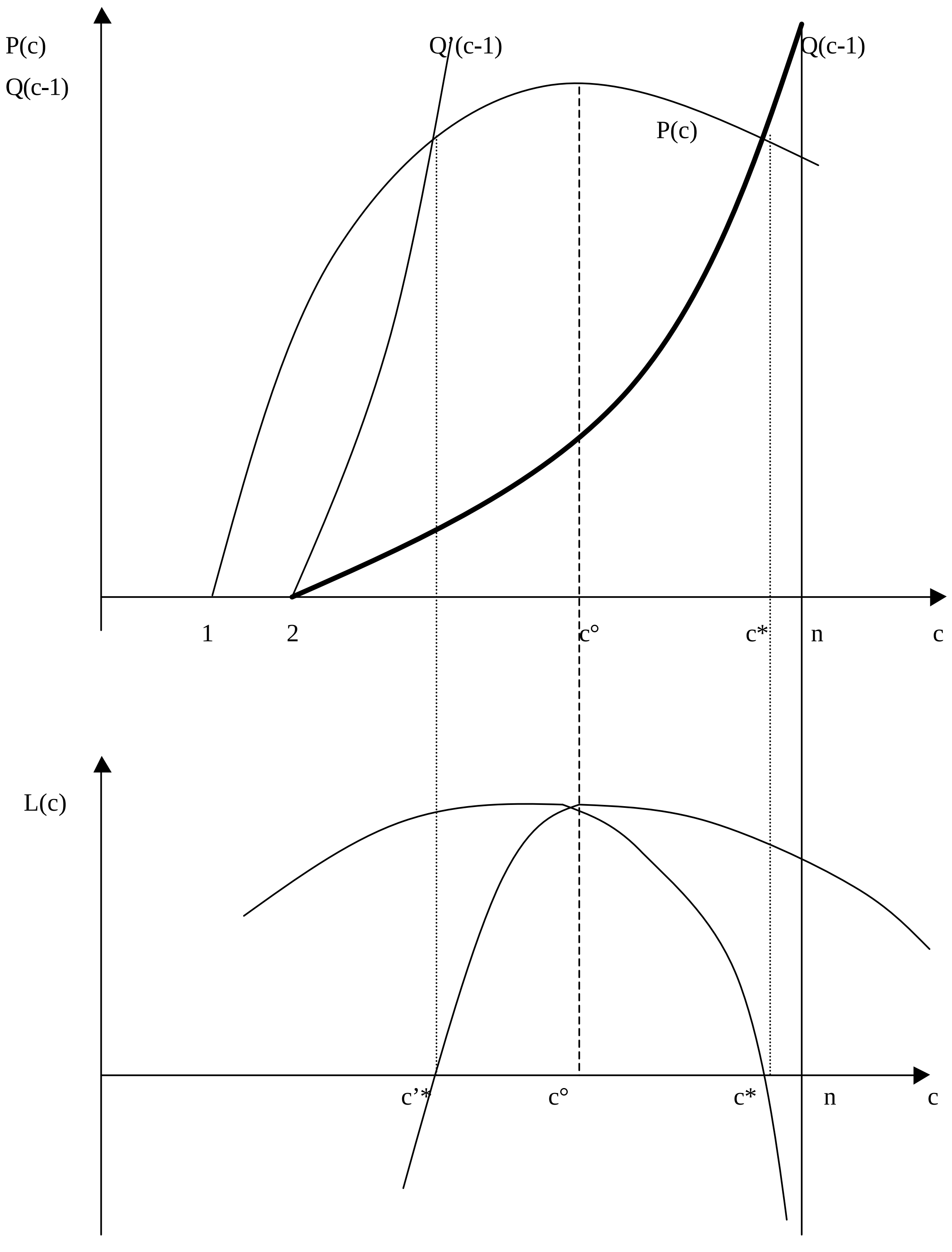


Figure 5. Stability function with non-orthogonal free-riding and monotonic payoff

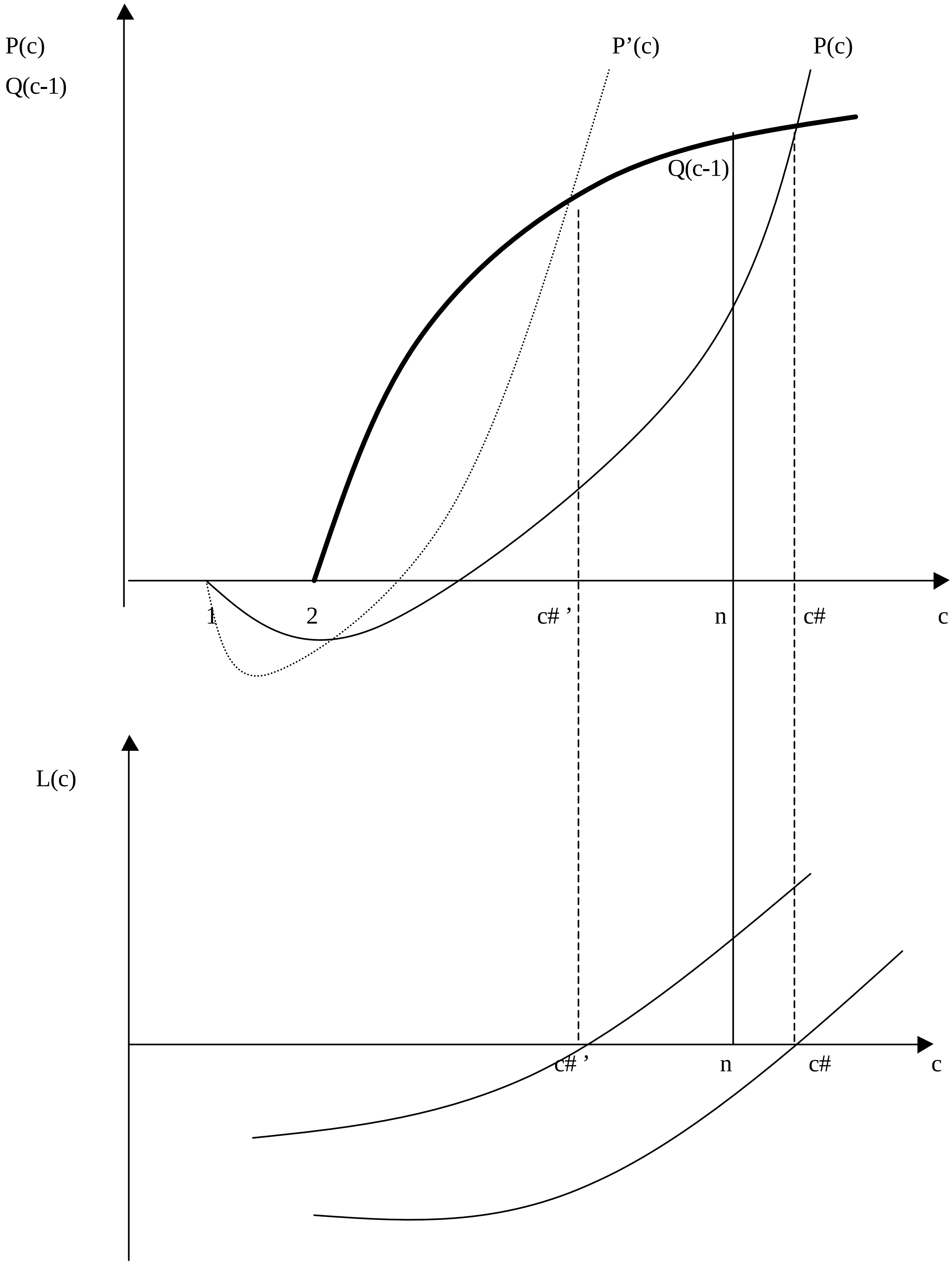
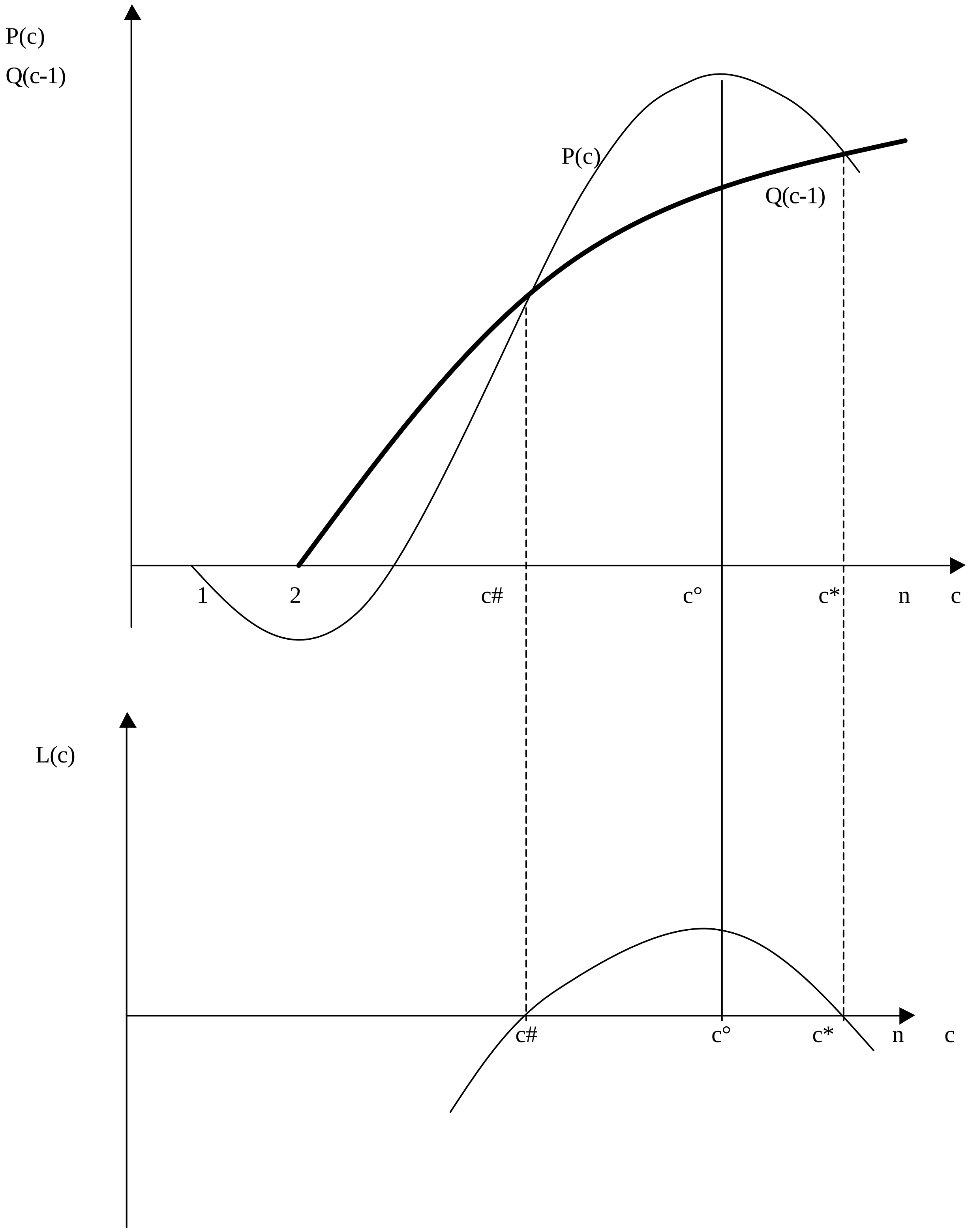


Figure 6. Stability function with non-orthogonal free-riding and humped-shaped payoff



In Figure 3, fixed abatement costs are implicitly assumed to be low. The presence of high fixed abatement costs to be shared among coalition members implies that the profitability function increases at decreasing rates (Heal, 1994). It is therefore concave, rather than convex. In this case (Figure 4), the intersection between the profitability and free-riding function can take place at values of c* close to n and even above n, thus implying that the grand coalition can be stable (P(n) > Q(n-1)).[14] This case has been proposed by Heal (1994) to show that a large coalition or even the grand coalition can emerge as the equilibrium outcome of the non-cooperative coalition game.

Stability function with non-orthogonal free-riding. In this case, cooperation may not be profitable because free-riders offset the cooperative effort of cooperators (Botteon and Carraro, 1997a). This is particularly true for small coalitions, when the number of free-riders is large and the number of cooperators is low. As a consequence, the function P(c) is first decreasing and then increasing and becomes positive for $c > c^m$ (see Figure 1). By contrast, the function Q(c-1) is always increasing but at decreasing rate. In environmental games, the economic mechanism (e.g. lower energy prices) behind the leakage effect usually exhibits decreasing returns, i.e. free-riders can damage cooperators at a decreasing rate, thus implying a concave, rather than convex, non-member partition function (see Figure 2). The resulting stability function L(c) is shown in Figures 5 and 6. In Figure 5, the profitability function is monotonic, in Figure 6 it is humped-shaped. Notice that, if P(c) and Q(c-1) intersect at $c^\# < n$, then a stable coalition may exist. However, $L(c^\#) \geq 0$, which implies that $c^\#$ does not satisfy the stability condition (2b), namely other players would like to join the coalition. Therefore, in the case of Figure 5, the grand coalition could be the stable one (see Proposition 2a below). In the case of Figure 6, the stable coalition under open membership is c* (see Proposition 2b). [15]

Let us now move to a formal analysis which enables us to prove the conclusions achieved by using the simple geometry of coalition stability. Let c* be the largest integer smaller than or equal to the value of c which satisfies:

(4a) $\quad L(c) = 0 \quad$ and $\quad L'(c) < 0.$

For convenience, let c*=1 when L(c) < 0 for all $1 < c \leq n$. Moreover, Let $c^\#$ be the largest integer smaller than or equal to the critical size c which satisfies:

(4b) $\quad L(c) = 0 \quad$ and $\quad L'(c) > 0$

[14] For this latter case to happen, some positive externalities within the coalition must exist. Otherwise, the likely outcome is a coalition c* larger than the coalition one would obtain without fixed costs but smaller than the grand coalition (Heal, 1994).

[15] If the profitability function is first convex and then concave, there may exist both $c^\#$ and c*, with $c^* > c^\#$ and $c^* < n$. In this case, the stable coalition would be the coalition c* rather than the grand coalition.

Proposition 1 (*Carraro and Siniscalco, 1993; Barrett, 1994*). In a game with Nash conjectures and open membership, in which the reaction functions are orthogonal, the stable coalition structures are:

- $\pi^* = \{c^*, 1_{(n-c^*)}\}$ when $1 < c^* < n$;
- the grand coalition structure $\pi^n = \{n\}$ when $c^* \geq n$, i.e. $L(c) > 0$ for all $c \in [2, n)$;

both when the per member partition function is monotonic and when it is humped-shaped (Figures 3 and 4).

Proof : Under orthogonal free-riding $c^m = 2$, because cooperators are not damaged by free-riders. Therefore, the profitability condition is satisfied for all coalitions $c \in [2, n]$. However, when $1 < c^* < n$, only c^* satisfies the stability conditions. Indeed, if c belongs to $(1, c^*)$, a singleton's best-reply is to join the coalition c. If c belongs to $(c^*, n]$, the best-reply of a cooperating player is to defect. Therefore, only when $c = c^*$, no player wants to change his/her strategy. Hence, the coalition c^* is stable. By contrast, the stable coalition structure of the game coincides with the grand coalition when $c^* \geq n$, i.e. $L(c) > 0$ for all $c \in [2,n)$. In this case, no player wants to join the coalition when $c = n$; for all c such that $1 < c < n$, a singleton's best-reply is always to join the coalition. Finally, notice that $L(c) < 0$ for all c belonging to $[2, n]$ cannot occur because both $P(c)$ and $Q(c-1)$ are monotonically increasing and $P(c)$ is defined in the interval $[1,n]$, whereas $Q(c-1)$ is defined in the interval $[2,n]$ (see Figures 3 and 4). Therefore, using Assumption 5, they can only intersect for $c > 2$. If the two curves do not intersect, $P(c) > Q(c-1)$, i.e. $L(c) > 0$; otherwise, there exists a value of c, say c', for which $L(c) = 0$ in the interval $[2,n]$. The stable coalition is defined by the largest integer smaller than or equal to c'. ¦

Proposition 2.a (*Carraro and Siniscalco, 1993)*: In a game with Nash conjectures and open membership, in which the profitability function is positive and monotonic for $c \geq c^m$, and free-riding is non-orthogonal (Figure 5), the stable coalition structures are:

- the singleton structure $\pi^S = \{1_n\}$ when $c^\# > n$.
- the grand coalition structure $\pi^n = \{n\}$ when $c^m \leq c^\# \leq n$.

Proof: In the presence of non orthogonal free-riding, by definition of c^m, the profitability condition is satisfied for all $c^m \leq c \leq n$. As for stability, when $c^\# > n$, the internal stability condition is not satisfied for all c belonging to $[2, n]$. Therefore, the stable coalition structure is $\pi^S = \{1_n\}$.
When $c^m \leq c^\# \leq n$ the internal stability condition is not satisfied for $c < c^\#$, whereas the external stability condition is not satisfied for all c in the interval $[c^\#, n)$, i.e. there is an incentive for free-rider to join a coalition c for all $c \geq c^\#$. Hence, no player wants to defect and no player wants to enter the coalition only when $c = n$. As a consequence, $\pi^n = \{n\}$ is the stable coalition structure. ¦

Proposition 2.b : In a game with Nash conjectures and open membership, in which the profitability function is positive for $c \geq c^m$ and humped-shaped, and free-riding is non-orthogonal (Figure 6), the stable coalition structures are:

- the singleton structure $\pi^S = \{1_n\}$ when $c^\# > n$.
- $\pi^* = \{c^*, 1_{(n-c^*)}\}$ when $c^m \leq c^\# \leq c^* \leq n$.

Proof : Same as for Proposition 2a. The only difference is that, when $c^m \leq c^\# \leq c^* \leq n$, there is no incentive to join a coalition $c > c^*$, i.e. when $c > c^*$, the best reply of a cooperating player is to defect. There is also an incentive to defect for all $c < c^\#$. External stability is not met for $c^\# \leq c < c^*$. Therefore, the stable coalition is c^*.

It is important to stress that the above results identify the stable coalition structures but do not describe the mechanisms through which these equilibria are reached. For example, to achieve the stable grand coalition in the case of non-orthogonal free-riding, an adequate minimum participation constraint may be necessary.[16] This minimum participation constraint coordinates the decision of all players in a way to provide all players with the incentive to join the coalition.

A simple numerical example can also help understanding the above analysis of stable coalitions. Table 2 shows the payoff that can be obtained by defectors and cooperators under different coalition structures and assuming orthogonal free-riding. In each row, the payoff of singletons is shown first. The values in Table 2 are computed using the simple model proposed in Carraro and Siniscalco (1992) and then used by Barrett (1994), Chander and Tulkens (1997) – even though Barrett (1994) adopts different rules of the game.

Let us first consider the structure of incentives that characterises this game. It can be summarised by the following three facts:

- *Fact 1*: the profitability condition is satisfied for all $c \geq 2$.
- *Fact 2*: the coalition is internally stable for all $c \leq 3$.
- *Fact 3*: the coalition is externally stable for all $c \geq 3$.

[16] A minimum participation constraint is a rule which says that a coalition forms only if at least a share α of the n players joins the coalition. A formal analysis of this case is in Carraro, Marchiori and Oreffice (2001) where the share α is endogenised and strategically determined by the n players. An adequate minimum participation constraint is a rule $\alpha^\#$ such that a coalition forms only if at least $c^\#$ players join it.

Table 2. A single coalition game with orthogonal free-riding

Coalition structure	Per-member partition function					
1,1,1,1,1,1,	0	0	0	0	0	0
1,1,1,1,2	2	2	2	2	0,5	0,5
1,1,1,3	6	6	6	2	2	2
1,1,4	12	12	4,5	4,5	4,5	4,5
1,5	20	8	8	8	8	8
6	12,5	12,5	12,5	12,5	12,5	12,5

As a consequence, the stable coalition structure of this game is $\pi^* = \{3,1,1,1\}$. Consider indeed any coalition structure in which the coalition is larger than three. By Fact 2, the best-reply of one of the coalition members is to defect from the coalition. On the other hand, consider any coalition structure in which the coalition is smaller than three. Then, by Fact 3 the best-reply of one of the players outside the coalition is joining the coalition. Fact 1 guarantees that the stable coalition is also profitable.

Now consider the case in which free-riding is non-orthogonal. The values of countries' payoff are summarised in Table 3. Again the incentive system is summarised by the following three facts:

- *Fact 1*: the profitability condition is satisfied for all $c \geq 5$
- *Fact 2*: the coalition is internally stable for all $c < 1$
- *Fact 3*: the coalition is externally stable for all $c \geq 1$

As a consequence, the stable coalition structure is $\pi^* = \{1,1,1,1,1,1\}$. Consider indeed any coalition structure in which the coalition is larger than one. By Fact 2, the best-reply of any coalition member is to defect until the singleton structure is reached. Furthermore, by Fact 3 no singleton would be better off by joining the coalition in any coalition structure.

Table 3. A single coalition game with non-orthogonal free-riding

Coalition structure	Per-member partition function					
1,1,1,1,1,1	1/49	1/49	1/49	1/49	1/49	1/49
1,1,1,1,2	1/36	1/36	1/36	1/36	1/72	1/72
1,1,1,3	1/25	1/25	1/25	1/75	1/75	1/75
1,1,4	1/16	1/16	1/64	1/64	1/64	1/64
1,5	1/9	1/45	1/45	1/45	1/45	1/45
6	1/24	1/24	1/24	1/24	1/24	1/24

What are the implications of the above results? First, the game structure which captures countries' interactions is not a prisoners' dilemma, but rather a chicken game, in which at least two groups of players (and two roles: signatories and defectors) co-exist (Carraro and Siniscalco, 1993). This has an important implication. If a stable coalition exists, the outcome of the game is not the one in which no cooperation takes place (no countries sign the agreement). At the equilibrium, there are two groups of countries, signatories and defectors, where the size of the group of signatories crucially depends on the slope of countries' reaction functions.

4.1.2 Rational conjectures

As well-known in oligopoly theory, the assumption of Nash conjectures is crucial in determining the feature of the equilibrium of the game. This is why several authors, starting from Bresnaham (1981), have proposed different assumptions on players' conjectures when they take their decision. Without entering into the debate on the validity of replacing the Nash assumption with alternative types of conjectures, we would like to consider the possibility, proposed for example in Chew (1994), Ray and Vohra (1999), Mariotti (1997), that a cooperator anticipates the reaction of the other players when leaving a coalition. For example, other cooperators may defect, or a free-rider may decide to join the coalition and replace the cooperator who defected. Similarly, a free-rider anticipates the reaction of the other players when he/she decides to join a coalition, etc. We assume that players' conjectures are rational, i.e. that each player correctly anticipates the reaction of the other players to his/her decision.

In this context, we need to re-define the equilibrium conditions of the game, i.e. the stability and profitability conditions. Whereas profitability is not modified by the introduction of rational conjectures, stability becomes as follows.

Rational conjecture stability (*Carraro and Moriconi,1998*). A coalition is stable iff it is internally and externally stable. It is internally stable if no cooperating player would be better off in the coalition structure induced by his defection. Formally:

(5a) $$p(c;\pi) \geq p(1;\pi')$$

for all players in the coalition c, where $\pi = \{c, 1_{(n-c)}\}$, $\pi' = \pi \backslash \{c\} \cup \{c', 1_{(k+1)}\}$ and $c' = c-k-1$, where k is the number of defections following i's defection (k+1 defections including i). It is externally stable if no singleton would be better off in the coalition structure induced by his accession to the coalition. Formally:

(5b) $$p(1;\pi) \geq p(c'';\pi'')$$

for all players which do not belong to c, where $\pi'' = \pi \backslash \{c, 1_{(k'+1)}\} \cup \{c''\}$, $c'' = c+k'+1$, and k' is the number of accessions following i's accession (k'+1 accessions including i).

Given this definition, and assuming the profitability condition to be met for simplicity's sake, the stable coalitions of the game are defined by the following proposition:

Proposition 3 (*Carraro and Moriconi, 1998*). In a game with rational conjectures and open membership, in which free-riding is orthogonal, the stable coalitions are:

- the grand coalition $c = n$, when $c^* \geq n$;
- a sequence of coalitions c^*, c^{**}, c^{***} … , where $c^{***} > c^{**} > c^*$ and $P(c^{**}) \geq Q(c^*)$, $P(c^{***}) \geq Q(c^{**})$, … , when $1 < c^* < n$, n is sufficiently large, and the per member partition function is either monotonic or humped-shaped.

Proof: When $c^* \geq n$, the equilibrium coalition of the game is the grand coalition. As $P(c) \geq Q(c-1)$ for all $c \leq n$, there is no incentive to defect, i.e. the internal stability condition is met for all $c \leq n$. In addition, there is an incentive to join the coalition for all $c \leq n$. Hence the stable coalition is $c = n$.

When $c^* < n$, and n is sufficiently large, we have a sequence of stable coalitions $c^* < c^{**} < c^{***}, \ldots, < n$. If $c < c^*$, a singleton's best-reply is always to join the coalition until c^* is reached. And no c^*+1 player wants to join c^*. Therefore, c^* is stable as in Proposition 1. However, other coalitions $c > c^*$ can be stable outcomes of the game. With rational conjectures, a cooperating player i anticipates the reaction of the others when leaving a coalition. In particular, a cooperating player of a coalition $c^{**} > c^*$ knows that if he/she defects from c^{**}, other k players will defect. If the conjecture is rational, player i knows that defections will continue until c^* is reached. Hence, the rational conjectures are $k=c^{**}-c^*-1$ and $k' = 0$. Player i compares $P(c^{**})$ -- the cooperative payoff in c^{**} -- with $Q(c^*)$ -- the free-riding payoff in the equilibrium coalition resulting from his/her deviation. If the former payoff is smaller than the latter, the cooperating player i defects. Otherwise, the coalition c^{**} is stable. Therefore, the coalition c^{**} is stable if $P(c^{**}) \geq Q(c^*)$ where $c^{**} = c^*+k+1$. When P(c) is monotonic, and if n is sufficiently large, it is always possible to identify a coalition size c^{**} such that $P(c^{**}) \geq Q(c^*)$. Therefore, a stable coalition c^{**} exists. Given c^{**}, and n sufficiently large, it is also possible to identify a coalition size c^{***} such that $P(c^{***}) \geq Q(c^{**})$. Hence, c^{***} is stable. And so on.

When P(c) is humped shaped, the above argument holds for values of c below the maximand of P(c) and if this maximand is large enough (which implies n large enough). ¦

Proposition 4.a: In a game with rational conjectures and open membership, in which the profitability function is positive and monotonic for $c \geq c^m$, and free-riding is non-orthogonal, the stable coalitions are:

- a sequence of coalitions $c^m, c^{**}, c^{***}\ldots$, where $c^m < c^{**} < c^{***} < \ldots$, and $P(c^m) \geq P(0)$, $P(c^{**}) \geq Q(c^m)$, $P(c^{***}) \geq Q(c^{**})$, … , when $c^\# > n$ and n is sufficiently large;
- when instead $c^m < c^\# \leq n$, the stable coalitions are a sequence of coalitions $c^m, c^{**}, c^{***} \ldots$ where $c^m < c^{**} < c^{***} < \ldots < c^\#$, and $P(c^m) \geq P(0)$, $P(c^{**}) \geq Q(c^m)$, $P(c^{***}) \geq Q(c^{**})$, … , and the grand coalition $c = n$.

Proof: If $c^\# > n$, no coalition $c \leq n$ is internally stable. Therefore the rational conjecture of players belonging to the coalition c is $k = c-1$ and $k' = 0$, i.e. the singleton coalition forms. The smallest coalition which satisfies the condition that the payoff of a cooperating player is larger than the payoff he/she would achieve when free-riding from the coalition resulting from his/her defection is c^m, because, by definition of c^m, $P(c^m) \geq P(0) = Q(0)$. Therefore, c^m is stable. However, if n is sufficiently large, another coalition size c^{**} exists such that $P(c^{**}) \geq Q(c^m)$; c^{**} is stable because a defector from c^{**} rationally conjectures $k = c^{**} - c^m - 1$ and $k' = 0$. Therefore when deciding whether or not to defect, he/she compares $P(c^{**})$ -- the payoff when cooperating in c^{**} -- with $Q(c^m)$ -- his/her payoff when free-riding on the coalition resulting from his/her defection. Since $P(c^{**}) \geq Q(c^m)$, c^{**} is stable. In a similar way, if n is sufficiently large, it is possible to show the stability of c^{***}, c^{****}, etc.

If $c^m < c^\# \leq n$ and $c^m < c^{**} < c^{***} < \ldots < c^\#$, again the rational conjectures of players in $c < c^\#$ is that the singleton structure would emerge at the equilibrium following a player's defection. Therefore, the previous argument holds again if $c^\#$ is large enough which implies n large enough. In addition, if a player belongs to c^{**} or c^{***} where $c^\# \leq c^{**} < c^{***} < \ldots \leq n$, his/her rational conjecture is that entry will follow until the grand coalition forms. i.e. k = 0 and k' = n-c-1. Therefore, given the monotonicity of the payoff function, the stable coalition is c = n. ¦

Proposition 4.b: In a game with rational conjectures and open membership, in which the profitability function is positive for $c \geq c^m$ and humped-shaped, and free-riding is non-orthogonal, the stable coalitions are:

- a sequence of coalitions c^m, c^{**}, $c^{***}\ldots$, where $c^m < c^{**} < c^{***} < \ldots$, and $P(c^m) \geq P(0)$, $P(c^{**}) \geq Q(c^m)$, $P(c^{***}) \geq Q(c^{**})$, … , when $c^\# > n$ and n is sufficiently large;
- when instead $c^m < c^\# < c^* \leq n$, the stable coalitions are a sequence of coalitions c^m, c^{**}, c^{***} … where $c^m < c^{**} < c^{***} < \ldots < c^\#$, and $P(c^m) \geq P(0)$, $P(c^{**}) \geq Q(c^m)$, $P(c^{***}) \geq Q(c^{**})$, … , and the coalition structure $\pi^* = \{c^*, 1_{(n-c^*)}\}$.

Proof: Same as for Proposition 4a. The only difference is that when $c^m < c^\# < c^* \leq n$ and $c > c^*$, the incentives for players in c are the same as those of Proposition 3. Hence, the same argument applies. ¦

The examples of Tables 2 and 3 can be useful to understand the implications of the introduction of rational conjectures in the coalition game. Consider first the case of orthogonal free-riding (Table 2). The starting point are again Facts 1 to 3. If the coalition structure contains a coalition c whose size is smaller than three, singletons have an incentive to join the coalition until the size-3 coalition is reached. As a consequence, the rational conjectures are k=0, k'=3-c+1 and the stable outcome is $\pi^* = \{1,1,1,3\}$. Once the structure $\{3,1,1,1\}$ is implemented, if a country defects, a singleton has an incentive to join the coalition. The rational conjectures are k=0, k'=1 and the equilibrium is again $\pi^* = \{1,1,1,3\}$.

If the coalition structure is $\pi' = \{1,1,4\}$, a coalition member is willing to defect. If a country defects the structure $\pi^* = \{1,1,1,3\}$ is reached. Since π^* is a stable coalition structure, it turns out that the rational conjectures are k=0 (no additional defections) and k'=0 (no accession to replace the defector). Hence, one of the cooperating countries, by comparing $p(4; \pi')$ and $p(1; \pi^*)$, decides to defect. Therefore, π' is not a stable structure (the stable structure is again π^*).

If the coalition structure is $\pi^{**} = \{1,5\}$, again a cooperating player has an incentive to defect. If he/she defects, the structure π' is reached; but this structure is not stable as previously shown. Hence, the rational conjectures are k = 1 (one more country will defect) and k' = 0 (no accession to replace the two defectors). Moreover, $p(5; \pi^{**}) > p(1;\pi^*)$. Therefore, by comparing $p(5; \pi^{**})$ with $p(1;\pi^*)$, no country belonging to c=5 decides to defect. As a consequence, the structure π^{**} is stable and a coalition c** = 5 forms.

If the coalition structure is $\pi'' = \{6\}$, a cooperating country is willing to defect. If it defects, the stable structure $\pi^{**} = \{1,5\}$ is reached. Then, the rational conjectures are k=0 (no further defection) and k'=0 (no accession to replace the defector). By comparing $p(6; \pi'')$ and $p(1; \pi^{**})$ a coalition member decides to defect. Hence, the grand coalition is not stable in this game.

The conclusion is therefore as follows. When players choose their strategies by adopting rational conjectures, the stable coalition structure of the game summarised by Table 1 is either the basic coalition structure $\pi^* = \{1,1,1,3\}$ or the coalition structure $\pi^{**} = \{1,5\}$.

Now consider Table 3, where the game with non-orthogonal free-riding is shown. If the coalition structure is such that the coalition size is smaller than 5, by Fact 2 no singleton wants to join the coalition. On the other hand, by Fact 1 each cooperating country has an incentive to defect and to implement the singleton structure. Hence, for all coalition structures containing a coalition c < 5 the rational conjectures are k=c-1 (all countries decide to defect) and k'=0 (no accession to replace the defectors). However, P(c) < P(0) for c < 5 by Fact 1 and no stable coalition forms, i.e. the stable coalition structure is the singleton structure.

If the coalition structure is $\pi^{**} = \{1,5\}$, each country belonging to c has an incentive to defect, but its defection would induce a coalition structure in which c < 5 and then the singleton coalition structure as argued above. Hence, in this case the rational conjectures are k = 5-1 (a defection from c would lead all other countries to defect) and k' = 0 (because no singleton has an incentive to join any coalition). This implies that each coalition member decides whether or not to cooperate by comparing his/her payoff $p(5; \pi^{**})$ with his/her payoff in the singleton structure $p(1; \pi^s)$. As a consequence, the coalition structure π^{**} is stable.

Finally, if the coalition structure is the grand coalition $\pi'' = \{6\}$, by Fact 1 a cooperating country has an incentive to defect. If it does defect, the coalition structure π' (which is stable as discussed above) is formed. Then, in this case the rational conjectures are k=0 (no further defection) and k'=0 (no accession to replace the defector). As a consequence, one of the cooperating countries, by comparing $p(6; \pi'')$ and $p(1; \pi^*)$, decides to defect. The grand coalition is not stable.

Therefore, in the case of Table 3 (non-orthogonal free-riding) with rational conjectures there are again two stable coalition structures: the trivial coalition structure π^s={1,1,1,1,1,1} and the structure π^* = {1,5}.

4.2 Exclusive membership

In the *exclusive membership game* (Yi and Shin, 1994) or *game* Δ (Hart and Kurz, 1983), each player can join a coalition only with the consensus of the existing members, but he/she is free to leave the coalition. In this decision process, each player's message consists in a list of players with whom he/she wants to form a coalition. Players who then announce the same list form a coalition, which is not necessarily formed by all players in the list. In order to capture the intuition which lies behind the exclusive membership game, the following *optimality criterion* is introduced:

Optimality. A coalition denies membership to an outside player only if this new accession is not profitable for the coalition. Hence, a coalition accepts a new member only if all its members would be better off in the larger coalition. A coalition is optimal when there is no incentive to accept additional members or to exclude existing members.

In order to determine the equilibrium coalitions under exclusive membership, we need to modify the stability condition which characterises the equilibrium of the game as follows:

Exclusive membership stability (*Yi and Shin, 1994*). A coalition is stable iff the two following conditions are satisfied:

- no cooperating player is better off by defecting (internal stability). Formally:

(6) $$p(c;\pi) \geq p(1;\pi'),$$

where $\pi=\{c,1_{(n-c)}\}$ and $\pi' = \pi\backslash\{c\}\cup\{c-1, 1\}$ for all $i\in c$.

- the coalition is optimal, i.e. no cooperating player has an incentive to increase or reduce the coalition size.

4.2.1 Nash Conjectures

The analysis focuses on the case in which the profitability function is humped-shaped. In the case of monotonic payoff functions, the results coincide with those obtained in the case of open membership, because the optimal coalition is the grand coalition in both cases.

Proposition 5: In a game with Nash conjectures and exclusive membership, in which the profitability function is humped-shaped and free-riding is orthogonal, the stable coalition structures are:

- $\pi^* = \{c^*, 1_{n-c^*}\}$, when $c^* < c°$;
- $\pi° = \{c°, 1_{n-c°}\}$, when $c^* \geq c°$.

where c° is the maximand of P(c).

Proof: First, consider the case $c^* < c°$. A coalition structure in which $c > c^*$ cannot be an equilibrium because the best-reply of a cooperating player is always to leave the coalition. A coalition structure in which $c < c^*$ is internally stable (eq. (6) is met), and no cooperating player has an incentive to ban new accessions. Therefore, the equilibrium of the game is $\pi^* = \{c^*, 1_{n-c^*}\}$. The coalition c* is the internally stable coalition which maximises the per-member payoff.
Consider the case $c^* \geq c°$. All coalitions $c' > c^*$ do not satisfy (6). All coalitions c' such that $c° < c' \leq c^*$ satisfy (6), but cooperating players agree that a smaller coalition would provide a larger payoff because $P(c') < P(c°)$ by definition of c°. All coalitions c" such that $c'' < c°$ satisfy (6), but cooperating players agree that a larger coalition would provide a larger payoff because $P(c'') < P(c°)$ by definition of c°. Therefore, the stable coalition structure is $\pi°=\{c°, 1_{n-c°}\}$. ¦

Let us use again a simple model with orthogonal free-riding to illustrate this result.[17] The simplest way to obtain a humped-shaped stability function is to introduce some coordination costs within the coalition and to assume that these costs increase with the coalition size.[18] The resulting payoff matrix is shown in Table 4.

The incentive structure is summarised by the following four facts:

- *Fact 1*: the profitability condition is satisfied for all c.
- *Fact 2*: the coalition is internally stable for all $c \leq 5$.
- *Fact 3*: the individually optimal coalition size is 4.

As a consequence, under open membership the equilibrium coalition of this game is $\pi^* = \{1,5\}$, whereas under exclusive membership the equilibrium coalition is $\pi° = \{1,1,4\}$ (we are in the case in which $c^* > c°$). Consider indeed any coalition structures in which the coalition is smaller than 5. By Fact 2, no cooperating player wants to leave the coalition. By Fact 4, coalition members prefer to refuse new cooperators once size four is reached. Hence, under the exclusive membership rule the size of the largest internally stable coalition

[17] The analysis of the example in which free-riding is non-orthogonal provides a marginal contribution to the comprehension of the result. It is therefore omitted.

[18] A less ad hoc approach, where positive and negative R&D externalities interact within the coalition, is proposed in Carraro and Siniscalco (1997).

is four. Now consider the structure with a coalition larger than five. By Fact 2, the best-reply of one of the cooperators is to defect. The grand coalition cannot be then a stable coalition structure.

Table 4. A single coalition game with humped-shaped payoff function and non orthogonal free-riding.

Coalition structures	Per-member partition function					
1,1,1,1,1,1,	0	0	0	0	0	0
1,1,1,1,2	0.7	0.7	0.7	0.7	2.7	2.7
1,1,1,3	2	2	2	4.4	4.4	4.4
1,1,4	4.2	4.2	5	5	5	5
1,5	6	4.4	4.4	4.4	4.4	4.4
6	2.7	2.7	2.7	2.7	2.7	2.7

Proposition 6: In a game with Nash conjectures and exclusive membership, in which the profitability function is positive for $c \geq c^m$ and humped-shaped, and free-riding is non-orthogonal, the stable coalition structures are:

- the singleton structure $\pi^S = \{1_n\}$ when $c^\# > n$;
- $\pi^* = \{c^*, 1_{n-c^*}\}$, when $c^m \leq c^\# \leq c^* \leq c°$;
- $\pi° = \{c°, 1_{n-c°}\}$, when $c^m \leq c^\# \leq c° < c^*$.

Proof : Same as for Proposition 2b. The only difference is that when $c^m \leq c^\# \leq c° < c^*$, the optimal coalition is no longer c*, but c°. ¦

4.2.2 Rational conjectures

Now consider the case in which players' conjectures are rational, i.e. each player correctly anticipates the reaction of the other players to his decision. The definition of stability must be modified to take into account both the introduction of rational conjectures and the adoption of exclusive membership .

Rational conjecture stability with exclusive membership. A coalition is stable iff it is internally stable and optimal. It is internally stable if no cooperating player would be better off in the coalition structure induced by his defection. Formally:

(7) $$p(c;\pi) \geq p(1; \pi')$$

for all players in the coalition c, where $\pi = \{c, 1_{(n-c)}\}$, $\pi' = \pi \backslash \{c\} \cup \{c', 1_{(k+1)}\}$ and $c'=c-k-1$, where k is the number of defections following i's defection (k+1 defections including i). It is optimal, if no cooperating player has an incentive to accept additional members or to exclude existing members.

Given this definition, the stable coalitions are identified by the following propositions (again, it makes sense to analyse only the case of humped shaped profitability functions):

Proposition 7 : In a game with rational conjectures and exclusive membership, in which the profitability function is humped shaped and free-riding is orthogonal, the stable coalition structures are:

- a sequence of coalitions c^*, c^{**}, c^{***} … , where $c^{***} > c^{**} > c^*$ and $P(c^{**}) \geq Q(c^*)$, $P(c^{***}) \geq Q(c^{**})$, … , when $1 \leq c^* < c^{**} < c^{***} < \ldots < c° < n$, and c° is sufficiently large;
- $\pi° = \{c°, 1_{n-c°}\}$, when $c° \leq c^* \leq n$ or $c^* > n$.

Proof: All coalitions c belonging to [2,n] are profitable. If $1 < c^* < c°$, all players outside a coalition $c < c^*$ have an incentive to join the coalition. Their rational conjectures are k=0 and $k'=c^*-c-1$. Therefore, c^* is stable. When $c^* < c < c°$, all players in c have an incentive to defect. However, if c° is sufficiently large, a coalition $c^{**} < c°$ may exist such that $c^{**} > c^*$ and $P(c^{**}) \geq Q(c^*)$. If this is the case, c^{**} is stable (see the proof of Proposition 3). Similarly, a coalition $c^{***} < c°$ may exist such that $c^{***} > c^{**}$ and $P(c^{***}) \geq Q(c^{**})$, etc. Whenever one of the coalitions c^{**}, c^{***}, etc., is larger than c°, the sequence stops.
If $c° < c^* < n$, all players outside a coalition c belonging to [c°, c*] have an incentive to join the coalition. However, c* does not meet the optimality condition because the optimal coalition is $c° < c^*$. The rational conjectures are k=0 and $k'=c°-c-1$. Therefore, c° is stable because players in c° have no incentive to defect (being $c° < c^*$). Finally, when $c° < c^* < n$, all players in a coalition $c > c^*$ have an incentive to defect. There could be a coalition c^{**} such that $P(c^{**}) \geq Q(c^*)$ which could be stable under the conjectures $k=c^{**}-c^*-1$ and k'=0. However, c^{**} does not meet the optimality condition, and therefore the rational conjectures are $k=c^{**}-c°-1$ and k'=0. However, even though $P(c^{**}) \geq Q(c°)$, c^{**} is not stable because it is not optimal ($c^{**} > c°$). Therefore, all coalition $c > c^*$ cannot be stable under exclusive membership. ¦

Proposition 8 : Consider a game with rational conjectures and exclusive membership, in which the profitability function is positive for $c \geq c^m$ and humped shaped, and free-riding is non-orthogonal. The stable coalition structures are:

- a sequence of coalitions c^m, c^{**}, $c^{***}\ldots$, where $c^m < c^{**} < c^{***} < \ldots < c°$, and $P(c^m) \geq P(0)$, $P(c^{**}) \geq Q(c^m)$, $P(c^{***}) \geq Q(c^{**})$, … , when: (i) $c^{\#} > n$ and c° is sufficiently large, or (ii) $c^m < c^{\#} \leq n$, $c^m < c^{**} < c^{***} < \ldots < c° < c^{\#}$ and c° is sufficiently large; (iii) $c^m < c^{\#} < c^* \leq c°$, $c^* < c^{**} < c^{***} < \ldots \leq c°$ and c° is sufficiently large.
- $\pi° = \{c°, 1_{n-c°}\}$, otherwise.

Proof: Same as for Propositions 4b and 7.

4.3 Coalition unanimity

Let us consider another membership rule often used in the economic theory of coalition stability, namely the so called coalition unanimity. In the *coalition unanimity game* (Yi and Shin, 1994; Chander and Tulkens, 1995; Bloch, 1997) or *game G* (Hart and Kurz, 1983), no coalition can form without the unanimous consensus of its members. This implies that players are not free to either join the coalition or to leave it. Therefore, this membership rule introduces restrictions both on entry (as the exclusive membership rule) and on exit behaviours of players. In the decision process, players' messages consist in a list of players as in the previous one. However, if a coalition is formed, it is necessarily composed of all players in the list and as soon as a player defects the coalition breaks up into singletons.

As underlined in Tulkens (1998), the coalition unanimity rule can be considered the opposite of the open membership rule. Under open membership, a player who defects assumes that players in the coalition keep cooperating after his/her defection. Instead, under coalition unanimity, a player knows that his/her decision to free-ride will lead the whole coalition to collapse.

Therefore, the coalition unanimity rule implicitly embodies a conjecture on players' behaviour after a defection takes place. Whereas open membership is characterised by the usual Nash conjecture, coalition unanimity assumes the following conjecture: a member of the coalition who decides to defect assumes that all other members of the coalition will follow his/her decision and will also defect. As said, after a defection the coalition breaks up into singletons. Therefore, a defecting player's payoff is the non-cooperative one.

As a consequence, the definition of stability must be modified as follows:

Coalition unanimous stability (*Chander and Tulkens, 1995*). A coalition is stable iff:

- the payoff of coalition members is larger than the payoff they would achieve in the singleton coalition structure π^S. Formally: $p(c; \pi) > p(1; \pi^S)$ for all players which belong to c, where $\pi = \{c, 1_{n-c}\}$, $\pi^s = \{1_n\}$.

- the coalition is optimal, i.e. no cooperating player has an incentive to accept additional members or to exclude existing members.

This definition of stability is less demanding than conditions previously provided. As a consequence, the equilibrium stable coalition is generally larger, as shown by the following propositions.

Proposition 9: In a game in which the profitability function is monotonic, the stable coalition structure under coalition unanimity is the grand coalition structure $\pi^n = \{n\}$ whatever the type of free-riding.

Proof : As previously seen, when free-riding is orthogonal, all coalitions $c \in [2, n]$ are profitable. Therefore $p(c; \pi) > p(1; \pi^S)$ for all $c \in [2, n]$. When the profitability function is monotonic, the optimal coalition is the grand coalition $c = n$. As a consequence, only the grand coalition satisfies the two conditions defining the coalition unanimous stability. When free-riding is non-orthogonal, all coalitions $c \in [c^m, n]$ are profitable, with $c^m \leq n$. Therefore $p(c; \pi) > p(1; \pi^S)$ for all $c \in [c^m, n]$. Again monotonicity implies the optimality of the grand coalition which thus satisfies the two conditions defining the coalition unanimous stability. ¦

Proposition 10 : In a game in which the profitability function is humped-shaped, the stable coalition structure under coalition unanimity is $\pi° = \{c°, 1_{(n-c°)}\}$ whatever the type of free-riding (when free-riding is non-orthogonal the condition $c^m < c°$ is necessary).

Proof : Again, when free-riding is orthogonal, all coalitions $c \in [2, n]$ are profitable. Therefore $p(c; \pi) > p(1; \pi^S)$ for all $c \in [2, n]$. When free-riding is non-orthogonal, all coalitions $c \in [c^m, n]$ are profitable, with $c^m \leq n$. When the profitability function is humped-shaped, the optimal coalition is c°, the maximand of P(c). When free-riding is orthogonal, obviously $c° \geq 2$. When free-riding is non-orthogonal, we need $c° \geq c^m$. As a consequence, if $c^m \leq c°$, only c° satisfies the two conditions defining the coalition unanimous stability. ¦

5. Sequential games

The simultaneity of decisions characterising the representation of the negotiating process discussed in the previous sections may be replaced by a sequential order of moves, in order to capture the idea of expectations on future reactions which was previously introduced in the case of rational conjectures and coalition unanimity. Therefore, we will focus only on the case of Nash conjectures in a sequential game, thus adopting the concept of subgame perfect Nash equilibrium (stability).

In this context, the stable coalition structures under open membership, exclusive membership and coalition unanimity will be identified, both for linear and circular games.

5.1. Linear sequential games

Consider first the case in which players decide one after the other, i.e. the order of moves is linear. Without loss of generality, let us assume that player 1 decides first, player two decides second and so on. Let us start by analysing the case of open membership. We have:

Proposition 11: If the profitability function is monotonic, and free-riding is orthogonal, under open membership the stable coalition structures are:

- $\pi^* = \{c^*, 1_{(n-c^*)}\}$ when $1 < c^* < n$;
- the grand coalition structure $\pi^n = \{n\}$ when $c^* \geq n$, i.e. $L(c) > 0$, for all $c \in [2,n)$;

Proof: First of all let us recall that, as in Proposition 1, the case $L(c) < 0$ for all $c \in [2,n]$ cannot occur. To identify the subgame perfect Nash equilibrium of the game consider again the definition of stability provided by (2a)(2b). This definition implies that player i, who decides at stage i, must choose whether or not to join the coalition c eventually formed in the previous stages of the sequential game. Player 1 decides in the first stage when no coalition exists. He/she therefore decides whether or not to be the first member of a coalition. Player i's best reply is a s follows: if at least c* players previously decided to form a coalition, his/her best option is to free-ride. Otherwise, player i decides to cooperate. In addition, by definition of c*, $Q(c^*) > P(c^*) \approx Q(c^*-1)$. Hence, all players know that at the equilibrium free-riders achieve a larger payoff than cooperators. Therefore, by anticipating the best reply of the following players, the first n-c* player decides to free ride, whereas the remaining c* player decides to cooperate (by definition of c*). As a consequence, c* is the stable coalition unless $L(c) > 0$ for all $c \in [2,n)$. In this latter case, the grand coalition forms. ¦

Proposition 12.a: If free-riding is non-orthogonal and the profitability function is positive and monotonic for $c \geq c^m$, under open membership the stable coalition structure of the linear sequential coalition game is the singleton structure $\pi^S = \{1_n\}$.

Proof: In the presence of non-orthogonal free-riding, by definition of c^m, the profitability condition is satisfied for all $c \in [c^m, n]$. As for stability, when $c^\# > n$, then no player has an incentive to join the coalition at any stage of the game. Therefore, the stable coalition structure is $\pi^S = \{1_n\}$. When $c^m \leq c^\# \leq n$, player i chooses to cooperate if at least $c^\#$ players previously decided to join the coalition. If not, he/she prefers to free-ride. Therefore, the first player chooses to free ride and the following ones take the same decision. The stable coalition structure is again $\pi^S = \{1_n\}$. ∎

Proposition 12.b: If free-riding is non-orthogonal and the profitability function is positive for $c \geq c^m$, but humped-shaped, under open membership the stable coalition structure of the linear sequential coalition game are:

- the singleton structure $\pi^S = \{1_n\}$ when $c^\# > n$,
- $\pi^* = \{c^*, 1_{(n-c^*)}\}$ when $c^m \leq c^\# \leq c^* \leq n$.

Proof : Again, by definition of c^m, the profitability condition is satisfied for all $c \in [c^m, n]$. Moreover, when $c^\# > n$, no player has an incentive to join the coalition at any stage of the game. Therefore, the stable coalition structure is $\pi^S = \{1_n\}$. When $c^m \leq c^\# \leq c^* \leq n$, player i's best reply is as follows. Player i decides to join the coalition if at least c' players, $c^\# \leq c' \leq c^*$, previously joined the coalition. If less than $c^\#$ players or more than c* players previously joined the coalition, player i prefers to free-ride. In addition, as already shown, $Q(c^*) > P(c^*) > Q(c^\#) > P(c^\#)$. Therefore, by anticipating the reaction of the other players, the first player chooses to free ride and the following ones take the same decision, until the last c* players decide to join the coalition. Therefore, the stable coalition structure is $\pi^* = \{c^*, 1_{(n-c^*)}\}$. ∎

In the case of exclusive membership and monotonic profitability function, the stable coalitions coincide with those obtained in the case of open membership. However, when the profitability function is humped-shaped, we have:

Proposition 13: If the profitability function is humped-shaped and free-riding is orthogonal, under exclusive membership the stable coalition structures of the linear sequential coalition game are:

- $\pi^* = \{c^*, 1_{(n-c^*)}\}$ when $1 < c^* < c^\circ$;
- $\pi^\circ = \{c^\circ, 1_{(n-c^\circ)}\}$ when $c^* \geq c^\circ$

Proof: Again, the case $L(c) < 0$, for all $c \in [2,n]$ cannot occur. In the case where $1 < c^* < c°$, the proof of the proposition coincides with the one of Proposition 11. When $c^* \geq c°$, there are three groups of players: the free-riders, the cooperators and those players who would like to cooperate but are excluded from the coalition. Given the definitions of c* and $c^* \geq c°$, we have $P(c°) > Q(c°)$ and $P(c°) > P(c^*) \approx Q(c^*-1)$. Therefore, when decisions are taken sequentially, the first c° players decide to cooperate, then c*-c° players are excluded from the coalition and forced to free-ride, and finally n-c* players decide to free-ride. As a consequence the stable coalition is c°. ¦

Finally, in the case of coalition unanimity:

Proposition 14: Under coalition unanimity, if the profitability function is positive for $c \geq c^m$ and monotonic, the stable coalition structure of the linear sequential coalition game is the grand coalition structure $\pi^n = \{n\}$ whatever the type of free-riding. If the profitability function is humped-shaped, the stable coalition structure becomes: $\pi° = \{c°, 1_{(n-c°)}\}$.

Proof: Under coalition unanimity, the first player to move knows that if he/she does not choose to join the coalition, no coalition will form. Since all coalitions c are profitable (either for $c\geq 2$ or for $c \geq c^m > 2$), then player 1 prefers to join the coalition. If the profitability function is monotonic, the same decision, and for the same reasons, will be taken by all the subsequent players. If the profitability function is humped-shaped, the first player to move knows that a coalition of size c° will form and that $P(c°) > Q(c°)$. Therefore, it is optimal for the first c° players to form a coalition, whereas the last n-c° players decides to free-ride. ¦

The main conclusion that can be derived from the above propositions is that the stable coalition structures of a linear sequential coalition game coincide, with a few exceptions, with the stable coalition structures of a simultaneous game. The main advantage of the sequential decision structure is that it enables us to identify which players play which role. In simultaneous games, the “identity” of cooperators is not determined. In sequential games, it is possible to know which players will join the coalition. For example, under open membership and orthogonal free-riding, the first n-c* players to move are the ones who free-ride, whereas the last c* movers are those who join the coalition. Under exclusive membership, the first c° players form a coalition, whereas the remaining ones free-ride (either because they are excluded from the coalition or because they are actual free-riders).

The examples provided by Tables 2 and 3 helps to understand our results. Consider Table 2 and the rule of open membership. From the decision tree of the game implicit in the payoff matrix, it is easy to see that the best-reply of the last player of the game is to defect if the coalition previously formed is larger or equal to three. On the other hand, if previous players formed a coalition smaller than 3, the last player’s best reply is to join the coalition. The same reasoning can be applied to all players. Hence, the first player decides to form a singleton, similarly the second player, etc. until the last three players form a coalition. The subgame perfect Nash equilibrium of this game is therefore $\pi^* = \{3,1,1,1\}$. In the case of non-orthogonal free-riding (Table 3) and open membership, the best-reply of each player at each node of the tree is to defect, whatever the size of the previously formed coalition. Hence, the subgame perfect Nash equilibrium structure is the one in which no coalition forms.

5.2 Circular sequential games

The last step of our analysis focuses on a further possible assumption on the order of moves of the game. Rather than assuming that players decide one after the other -- from the first to the last one – we assume that after the last player has decided, the first player can revise his/her decision. Then the second one can revise his/her decision, and so forth. Therefore, the negotiation goes on sequentially until no player wants to change his/her choice.

The game consists of a sequence of rounds and each round is a linear sequential game. For each player, there is always a next round unless all players stick to the choice made in the previous round. This gets rid of the strategic disadvantage of being the last players in the game. Moreover, either before defecting or before joining the coalition, each player takes into account the reaction of all the other players in the game. Indeed, in a circular sequential game, each player chooses as if he was the first player of a linear sequential game.

Let us focus first on the stable coalition structures of a circular sequential game under open membership. We restrict our analysis to the case of orthogonal free-riding (where the profitability function is positive for all $2 \leq c \leq n$). The extension to the case of non orthogonal free-rider is trivial and is left to the reader.

Proposition 15. In a sequential game in which the order of moves is circular, the profitability function is monotonic, and free-riding is orthogonal, the stable coalition structure of the game under open membership is the grand coalition $\pi^n = \{n\}$, if: (i) $p(n; \pi^n) \geq p(1; \pi^*)$ and $1 < c^* < n$; or (ii) $c^* \geq n$; where $p(n; \pi^n)$ is a player's payoff when the grand coalition forms and $p(1; \pi^*)$ is the payoff as a free-rider in the structure $\pi^*= \{c^*, 1_{(n-c^*)}\}$.

Proof: Assume $1 < c^* < n$ and consider the behaviour of the first player. If he/she chooses to defect, he/she knows that at most the last c^* player will choose to cooperate when $1 < c^* < n$. Therefore, he/she gets a payoff equal at most equal to $p(1; \pi^*)$. If he/she chooses to cooperate and one of the subsequent players free-rides, the first player can always revise his/her decision and choose not to join the coalition, thus obtaining again at most $p(1; \pi^*)$. However, if he/she chooses to cooperate and all subsequent players join the coalition, he/she gets the grand coalition payoff $p(n; \pi^n)$. Therefore, if $p(n; \pi^n) \geq p(1; \pi^*)$, the first player can always take the risk to announce that he/she will join the coalition, because if all other players join the coalition, he/she gets $p(n; \pi^n) \geq p(1; \pi^*)$; otherwise, he she gets at most $p(1; \pi^*)$. Therefore, in the first round the first player announces that he/she joins the coalition. In a circular game, the second player behave as if he was the first player, and therefore he/she chooses according to the same reasoning. As a consequence, the second player also chooses to join the coalition. Similarly, the third player, the fourth one, and so forth. At the end of the first round, all players have chosen to join the coalition and the grand coalition forms. Therefore, the stable coalition structure is $\pi^n = \{n\}$. In the case $c^* \geq n$, all players have of course an incentive to join the coalition and the grand coalition forms. ¦

The next step to consider is the case in which the membership rule is exclusive. As in the previous sections, this case does not coincide with the one in which the membership rule is open when the per member partition function is humped-shaped. Then:

Proposition 16. In a sequential game in which the order of moves is circular, the profitability function is humped-shaped, and free-riding is orthogonal, the stable coalition structure of the game under exclusive membership is $\pi° = \{ c°, 1_{(n-c°)}\}$ if: (i) $p(c°; \pi°) \geq p(1; \pi^*)$ and $c^* < c°$ or (ii) $c° \leq c^*$.

Proof: Same as for the previous proposition. The only difference is that players find it optimal not to form a coalition larger than c°. ¦

The final case to be considered is the one in which the membership rule is coalition unanimity.

Proposition 17. In a sequential game in which the order of moves is circular, the profitability function is monotonic, and free-riding is orthogonal, the stable coalition structure of the game under coalition unanimity is the grand coalition $\pi^n = \{n\}$. If the profitability function is humped-shaped, the stable coalition structure is $\pi° = \{ c°, 1_{(n-c°)}\}$.

Proof: Same as for Proposition 15. The only difference is that, when the profitability function is monotonic, each player compares $p(n; \pi^n)$ with $p(1; \pi^S)$ because a defection induces all other players to defect. In the case of orthogonal free-riding, all coalitions $c \in [2, n]$ are profitable and therefore the condition $p(n; \pi^n) \geq p(1; \pi^S)$ is satisfied. As a consequence, the first player announces that he/she joins the coalition in the first stage and all other players follow. In the case of humped-shaped profitability function, the comparison is between $p(c°; \pi°)$ and $p(1; \pi^S)$, where $p(c°; \pi°) \geq p(1; \pi^S)$ because all coalitions $c \in [2, n]$ are profitable. ¦

Again consider the examples of Tables 2 and 3. In the case of orthogonal free-riding (Table 2), the equilibrium structure is the grand coalition. Consider any player in the grand coalition. By Fact 2, he/she has an incentive to defect. If he/she defects, the next player defects as well. Then, a third player deviates and so on. The process stops at $\pi^*=\{1,1,1,3\}$, because in this last structure there is no further incentives to defect. When the second round starts, the first player, in order to decide his strategy, compares $p(6; \pi^n) = 12.5$ -- his grand coalition payoff -- with $p(1; \pi^*) = 6$ -- the payoff he/she obtains as a consequence of his/her defection. Therefore, this player has an incentive to revise his decision and to choose to join the coalition. But if he/she would be better off by cooperating in the second round, he/she chooses to join the coalition already in the first round. The same reasoning can be applied to all players, thus showing that all players choose to join the coalition.

A similar conclusion can be derived by examining Table 3, where free-riding is non-orthogonal. In this case, the cooperative payoff is 1/24 which is larger than 1/49 -- the payoff a player would obtain by defecting from the grand coalition, after his/her defection is followed by all other players' defection. Therefore, in the first round all players choose to join the coalition.

Summing up, the possibility of revising their choice gives players the possibility to take the risk of joining the coalition in the first round. Ex-post this strategy is rewarded, because all players join the coalition, unless they prefer to free-ride on the partial stable coalition c* that may form at the equilibrium. Therefore, an agreement in which countries can revise their choice is not "weaker" than an agreement in which signatories are committed to their decision. Indeed, the possibility of revision enhances the incentives for players to sign the agreement (i.e. to join the coalition).

6. Conclusions

The above analyses have identified the stable coalition structures of different games. The common denominator of all games is that in the first stage all players choose whether or not to join a coalition and then in the second stage they set their policy variables. However, the games analysed in the previous sections differ with respect to other important features: the membership rules, the order of moves, the players' conjectures, the shape of the profitability function, the type of free-riding. By changing these features of the game, the stable coalition changes and multiple stable coalitions can emerge.

Table 2 and 3 below summarises part of our results.[19] Even when assuming Nash conjectures, it is clear that coalitions of different sizes can become stable by slightly modifying the rules and features of the game.

[19] Please refer to the Propositions presented in Section 4 and 5 for the exact conditions upon which relies the validity of the results summarised by Tables 2 and 3.

Stable coalition structures include the no coalition case, partial coalitions and the grand coalition. Does this lead to the conclusion that coalition theory is of no help in predicting the stable outcomes of a coalition game? The answer is negative. We would rather conclude that coalition theory suggests a set of rules which make it easier to achieve a large and stable coalition. For example, coalition unanimity, if adopted by the negotiating players, is of great help in achieving a large coalition (even the grand coalition). This is true in the case of symmetric players but it may not be true when players are asymmetric. Most importantly, a coalition unanimity rule may not be adopted by the players, even if it leads to the social efficient outcome, because players may have an incentive to free-ride in the game in which the membership rule is adopted.

This result is shown in Carraro, Marchiori and Oreffice (2001) where the choice of the membership rule is endogenised. The game is therefore formed by three stages. In the first one, players agree on the membership rule, is the second one, they decide whether or not to join the coalition, and in the third one they set their policy variables. Carraro, Marchiori and Oreffice (2001) show that coalition unanimity may not be an equilibrium of the game, even though players generally "tight their hands" by choosing a minimum participation constraint at the equilibrium.

Therefore, the conclusion of this paper that coalition unanimity, circular sequential moves and orthogonal free-riding are features of the game which favour the stability of large coalitions is not of great help unless more work is undertaken to establish whether and under what conditions these features of the game can be adopted or implemented by the players of the game. For example, in an international environmental negotiations, can players agree ex ante that, whatever the outcome of the negotiations, free-riding agents will not reduce but will also not increase their emissions? Can they agree ex ante to decide sequentially with the possibility of revising their choice ? Are there the incentives to agree on this type of choices (in many actual cases, under unanimity adoption of these ex-ante rules)?

The answer to these questions requires further research on the theory of coalition stability. This new research direction will enhance the applicability of coalition theory to many economic issues and policy relevant problems.

Table 2. Stable coalitions with orthogonal free-riding and Nash conjectures

	Open membership (monotonic or humped-shaped payoff)	Exclusive membership (humped-shaped payoff)	Coalition unanimity (monotonic or humped-shaped payoff)
Simultaneous game	$\pi^*=\{c^*,1_{n-c^*}\}$,when $1<c^*<n$ $\pi^n=\{n\}$, when $c^*\geq n$	$\pi^*=\{c^*,1_{n-c^*}\}$, when $1<c^*<c^\circ$ $\pi^\circ=\{c^\circ,1_{n-c^\circ}\}$, when $c^*\geq c^\circ$	*Monotonic payoff:* $\pi^n=\{n\}$ *Humped-shaped payoff:* $\pi^\circ=\{c^\circ,1_{n-c^\circ}\}$
Linear sequential game	$\pi^*=\{c^*,1_{n-c^*}\}$,when $1<c^*<n$ $\pi^n=\{n\}$, when $c^*\geq n$	$\pi^*=\{c^*,1_{n-c^*}\}$,when $1<c^*<c^\circ$ $\pi^\circ=\{c^\circ,1_{n-c^\circ}\}$, when $c^*\geq c^\circ$	*Monotonic payoff:* $\pi^n=\{n\}$ *Humped-shaped payoff:* $\pi^\circ=\{c^\circ,1_{n-c^\circ}\}$
Circular sequential game	$\pi^n=\{n\}$	$\pi^\circ=\{c^\circ,1_{n-c^\circ}\}$	*Monotonic payoff:* $\pi^n=\{n\}$ *Humped-shaped payoff:* $\pi^\circ=\{c^\circ,1_{n-c^\circ}\}$

Table 3. Stable coalitions with non-orthogonal free-riding and Nash conjectures

	Open membership (monotonic or humped-shaped payoff)	Exclusive membership (humped-shaped payoff)	Coalition unanimity (monotonic or humped-shaped payoff)
Simultaneous game	*Monotonic payoff:* $\pi^S = \{1_n\}$, when $c\# > n$ $\pi^n = \{n\}$, when $c^m < c\# \leq n$ *Humped-shaped payoff:* $\pi^S = \{1_n\}$, when $c\# > n$ $\pi^* = \{c^*, 1_{n-c^*}\}, c^m \leq c\# \leq c^* \leq n$	$\pi^S = \{1_n\}$, when $c\# > n$ $\pi^* = \{c^*, 1_{n-c^*}\}, c^m \leq c\# \leq c^* \leq c°$ $\pi° = \{c°, 1_{n-c°}\}, c^m \leq c\# \leq c° \leq c^*$	*Monotonic payoff:* $\pi^n = \{n\}$ *Humped-shaped payoff* $\pi° = \{c°, 1_{n-c°}\}$
Linear sequential game	*Monotonic payoff:* $\pi^S = \{1_n\}$ *Humped-shaped payoff:* $\pi^S = \{1_n\}$, $c\# > n$ $\pi^* = \{c^*, 1_{n-c^*}\}$, $c^m \leq c\# \leq c^* \leq n$	$\pi^S = \{1_n\}$, when $c\# > n$ $\pi^* = \{c^*, 1_{n-c^*}\}, c^m \leq c\# \leq c^* \leq c°$ $\pi° = \{c°, 1_{n-c°}\}, c^m \leq c\# \leq c° \leq c^*$	*Monotonic payoff:* $\pi^n = \{n\}$ *Humped-shaped payoff:* $\pi° = \{c°, 1_{n-c°}\}$
Circular sequential game	$\pi^n = \{n\}$	$\pi° = \{c°, 1_{n-c°}\}$	*Monotonic payoff:* $\pi^n = \{n\}$ *Humped-shaped payoff:* $\pi° = \{c°, 1_{n-c°}\}$

References

Barrett, S. (1994), “Self-Enforcing International Environmental Agreements”, Oxford Economic Papers, 46, 878-894.

Bloch, F. (1996), “Sequential Formation of Coalition in Games with Externalities and Fixed Payoff Division”, Games and Economic Behaviour, 14, 90-123.

Bloch, F. (1997), “Noncooperative Models of Coalition Formation in Games with Spillovers”, in C. Carraro and D. Siniscalco, eds., New Directions in the Economic Theory of the Environment, Cambridge University Press, Cambridge.

Botteon, M. and C. Carraro (1997a) "Burden-Sharing and Coalition Stability in Environmental Negotiations with Asymmetric Countries", in C. Carraro, ed., International Environmental Negotiations: Strategic Policy Issues, Edward Elgar, Cheltenham.

Botteon, M. and C. Carraro (1997b) "Environmental Coalitions with Heterogeneous Countries: Burden Sharing and Carbon Leakage", in A. Ulph, ed., Environmental Policy, International Agreements and International Trade, Oxford, Oxford University Press.

Bresnaham, T. (1981), “Duopoly Models with Consistent Conjectures”, American Economic Review, 71, 934-945.

Carraro, C. (1998), “Beyond Kyoto: A Game Theoretic Perspective”, in the Proceedings of the OECD Workshop on “Climate Change and Economic Modelling. Background Analysis for the Kyoto Protocol”, Paris, 17-18.9, 1998.

Carraro, C. (1999), “The Structure of International Agreements on Climate Change”, in C. Carraro (ed.), International Environmental Agreements on Climate Change, Kluwer Academic Publishers, Dordrecht.

Carraro, C. (2000), “Roads Towards International Environmental Agreements”, in H. Siebert (ed.), The Economics of International Environmental Problems, Mohr Siebeck, Tubingen.

Carraro, C., Marchiori, C. and S. Oreffice (2001), “Endogenous Minimum Participation in International Environmental Treaties”, mimeo, FEEM, Venice.

Carraro, C. and F. Moriconi (1998), “International Games on Climate Change Control”, FEEM Working Paper 56.98, FEEM, Milan.

Carraro, C. and D. Siniscalco (1992), “The International Protection of the Environment: Voluntary Agreements among Sovereign Countries”, in P. Dasgupta and K.G. Mäler (eds.), The Economics of Transnational Commons, Clarendom Press, Oxford.

Carraro, C. and D. Siniscalco (1993), “Strategies for the International Protection of the Environment”, Journal of Public Economics, 52, 309-328.

Carraro, C. and D. Siniscalco (1997), “R&D Cooperation and the Stability of International Environmental Agreements”, in C. Carraro, ed., International Environmental Agreements: Strategic Policy Issues, E. Elgar, Cheltenham.

Chander, P. and H. Tulkens (1993), “Strategically Stable Cost-Sharing in an Economic-Ecological Negotiation Process”, forthcoming in K.G. Mäler, ed., International Environmental Problems: An Economic Perspective, Kluwer Academic Publishers.

Chander, P. and H. Tulkens (1995), “A Core-Theoretical Solution for the Design of Cooperative Agreements on Trans-Frontier Pollution”, FEEM Working Paper 22.95, FEEM, Milan.

Chander, P. and H. Tulkens (1997), “The Core of an Economy with Multilateral Environmental Externalities”, International Journal of Game Theory, vol. 26, pp. 379-401.

Chwe, M.S.Y. (1994), “Farsighted Coalition Stability”, Journal of Economic Theory, 54, 299-325.

D’Aspremont, C.A. and J.J. Gabszewicz (1986), “On the Stability of Collusion”, in G.F. Matthewson and J.E. Stiglitz, eds., New Developments in the Analysis of Market Structure, Mac Millan Press, New York, 243-264.

D’Aspremont, C.A., A. Jacquemin, J.J. Gabszewicz and J. Weymark (1983), “On the Stability of Collusive Price Leadership”, Canadian Journal of Economics, 16, 17-25.

Donsimoni, M.P., N.S. Economides and H.M. Polemarchakis (1986), “Stable Cartels”, International Economic Review, 27, 317-327.

Finus M. and B. Rundshagen (2001), “Endogenous Coalition Formation in Global Pollution Control”, mimeo.

Hart, S. and M. Kurz (1983), “Endogenous Formation of Coalitions”, Econometrica, 51, 1047-1064.

Heal, G. (1994), "The Formation of Environmental Coalitions", in Carraro, C., ed., Trade, Innovation, Environment, Kluwer Academic Publisher, Dordrecht.

Hoel, M. (1991), "Global Environmental Problems: The Effects of Unilateral Actions Taken by One Country", Journal of Environmental Economics and Management, 20, 1, 55-70.

Konishi, H., Le Breton, M. and S. Weber (1997), “Stable Coalition Structures for the Provision of Public Goods”, in C. Carraro and D. Siniscalco, eds., New Directions in the Economic Theory of the Environment, Cambridge University Press.

Le Breton, M. and S. Weber (1993), "Stability of Coalition Structures and the Principle of Optimal Partitioning", mimeo, GREQE, Aix-Marseille.

Mariotti, M. (1997), “A Model of Agreements in Strategic Form Games”, Journal of Economic Theory, 73.

Ray, D. and R. Vohra (1999), “A Theory of Endogenous Coalition Structure”, Games and Economic Behavior, vol. 26, pp. 286-336.

Tulkens, H. (1998), "Cooperation vs. Free-Riding in International Affairs: Two Approaches", in H. Folmer and N. Hanley, eds., Game Theory and the Global Environment, E. Elgar, Cheltenham.

Ulph, A. (1993), "Environmental Policy, Plant Location, and Government Protection", FEEM Working Paper 43-93, FEEM, Milan.

Yi, S. (1997), "Stable Coalition Structures with Externalities", Games and Economic Behaviour, 20, 201-223.

Yi, S. and Shin, H.S. (1994), "Endogenous Formation of Coalitions in Oligopoly: 1. Theory", Dartmouth College, Working Paper 94-18.

Yi, S. and Shin, H.S. (2000), "Endogenous Formation of Research Coalitions with Spillovers", International Journal of Industrial Organization, vol. 18, pp. 229-256.