# International Games on Climate Change Control

by

Carlo Carraro
(University of Venice)

and

Francesca Moriconi
(Fondazione Eni E.Mattei)

*Abstract*

In this paper a n-player non-cooperative game is used to model countries' decision of whether or not to sign an international agreement on climate change control. The stable coalition structure of the game is defined and then computed for a climate game in which the role of carbon leakage is also taken into account. At the equilibrium, a coalition may emerge despite the public good nature of climate. The size of the coalition depends on the degree of interdependence of countries' emission strategies, and on the type of conjectures that each country forms on the consequences of its own decision to join or to defect from the environmental coalition.

Paper prepared for the special issue of *Natural Resource Modelling* on "Environmental Games". The Authors are grateful to Domenico Siniscalco and Henry Tulkens for helpful discussions.

First draft: September 1997.

Correspondence Address: Prof. Carlo Carraro, Department of Economics, University of Venice, S. Giobbe 873, Venice, Italy. Tel: +39 41 2574166, Fax: +39 41 2574176, E-mail: ccarraro@unive.it.

# International Games on Climate Change Control

## 1. Introduction

Climate change is becoming one of the major environmental concern of most countries in the world. Negotiations to control potential climate changes have been taking place, from Rio to Kyoto, for the last five years. There is a widespread consciousness that the risk of incurring in relevant economic and environmental losses due to climate change is high. Scientific analyses have become more and more precise on the likely impacts of climate change. According to the Second Assessment Report of the Intergovernmental Panel on Climate Change, current trends in GHG emissions may indeed cause the average global temperature to increase by 1-3,5 degrees C over the next 100 years. As a result, sea levels are expected to rise by 15 to 95 cm and climate zones to shift towards the poles by 150 to 550 km in mid latitudes. In order to mitigate the adverse effects of climate change, the IPCC report concludes that a stabilisation of atmospheric concentration of carbon dioxide -- one of the major GHG -- at 550 parts per million by volume (ppmv) is recommended. This would imply a reduction of global emissions of about 50 per cent with respect to current levels.

In this context, countries are negotiating to achieve a world-wide agreement on GHG emissions control in order to stabilise climate changes. There are a few distinguishing features of these negotiations:

- all world countries are involved and required to take a decision on whether or not to sign a protocol with important implications on their energy -- and economic -- policies;
- no supra-national authority can enforce such a protocol which must therefore be signed on a voluntary basis;
- no commitment to cooperation is likely to be credible. Only positive economic net benefits, which include environmental benefits, can lead countries to adhere to an international agreement on climate change control;
- climate is a public good. As a consequence all countries are going to benefit from the action taken by a subgroup of one or more countries. There is therefore a strong incentive to free-ride;
- parties involved in the negotiations seem to be conscious that an agreement signed by all world countries is not likely and that the effort of GHG emission abatement has to be concentrated on a sub-group of (more developed ) countries. Current negotiations aim at defining "quantified legally-binding objectives for emission limitations and significant overall reductions within specified time frames" for Annex I countries[1].

[1] Annex I countries include the non-EU OECD countries, the European Union, Central and Eastern European countries, and newly independent states.

The goal of this paper is the analysis of the incentives that countries have to sign an international agreement on climate change control. Starting from the basic features of the climate change problem, we analyse under what conditions on the number of countries, the damaging effects of free-riding, the structure of costs and benefits, a coalition, i.e. a group of signatories of the international agreement, can emerge. The paper also clarifies the importance of the assumptions on the coalition behaviour, and on the conjectures that countries form on the consequences of their own decision to defect from a coalition (or to join it).

The paper shows that, despite the public good nature of climate, a coalition may emerge endogenously, i.e. a group of countries, but not all countries, may have the incentive to sign the agreement. The size of this group depends on the status quo (the initial coalition), the degree of interdependence of countries' emission strategies, and the conjectures formed on the implications of each country's action.

The structure of the paper is as follows. In Section 2 we provide the necessary definitions and assumptions. Section 3 determines the equilibrium of the environmental game and analyses its implications. Section 4 extends the results to two cases in which either the coalition is designed according to the "coalition unanimity rule" or countries are assumed to form "rational conjectures" on the consequences of their own strategy. A concluding section proposes directions for further research.

## 2. Assumptions and Definitions

Assume negotiations take place among n countries, $n \geq 3$, each indexed by $i=1, ..., n$. Countries facing an international environmental problem 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 of cooperating countries). In the second stage, they play the non cooperative Nash emission game, where the countries which signed the agreement play as a single player and divide the resulting payoff according to a given burden-sharing rule (any of the rules derived from cooperative game theory).[2]

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.

[2] This approach has to be contrasted with the traditional cooperative game approach (e.g. Chander-Tulkens, 1993, 1995) and with a repeated game approach (Barrett, 1994, 1997b). Moreover, notice that the regulatory approach often proposed in public economics is not appropriate given the lack of a supranational authority.

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. $C_i \cap C_j = \emptyset$ for $i \neq j$ and $\bigcup_{i=1}^{m} C_i = N$.

Since the formation of a coalition creates externalities, the appropriate framework to deal with this game is a *game in partition function form*, in which the payoff of each player depends on the entire coalition structure to which he belongs (Bloch, 1997; Ray and Vohra, 1996). This is why we convert the game in normal form into a game in partition function form.

Denote by $\Pi$ the set of all feasible coalition structures. A *partition function* P: $\Pi \rightarrow R$ is a mapping which associates to each coalition structure $\pi$ a vector in $\Re^{|\pi|}$, representing the worth of all coalitions in $\pi$. In particular, $P(C_i; \pi)$ assigns to each coalition $C_i$ in a coalition structure $\pi$ 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 $\pi$ a vector of individual payoffs in $\Re^n$; $p(C_i; \pi)$ represents the payoff of a player belonging to the coalition $C_i$ in the coalition structure $\pi$ [3].

Under suitable assumptions (A.2 and A.3 below), the second stage of the game can be reduced to a partition function (Yi, 1997; Bloch 1997). Therefore, the study of coalition formation consists of the study of the first stage of the game, i.e. the negotiation process between the players. This negotiation process can be modelled either as a simultaneous game, in which all the players announce at the same time their strategic choice, or as a sequential one, in which each player can announce his strategy according to an exogenous rule of order. Here we assume that:

A.1. All countries decide <u>simultaneously</u> in both stages;[4]
A.2. The second stage emission game has a <u>unique</u> Nash equilibrium for any coalition structure.

This assumption is necessary for the second stage of the game to be reduced to a partition function. However, in order to convert the strategic form into a partition function, the competition among the various coalitions has also to be specified. The common and perhaps the most natural assumption is that:

---

[3] Bloch (1997) denotes the per-member partition function by the term "valuation".

[4] By contrast, Barrett (1994) assumes that the group of signatories is Stackelberg leader with respect to non-signatories in the second stage emission game. In Bloch (1997) it is assumed that countries play sequentially in the first stage coalition game.

A.3. 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 $\pi = \{C_1, C_2, \ldots, C_m\}$, let x* be a vector of strategies such that

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

Then define:

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

Studying the issue of coalition formation by partition function games implies a limitation. The second-stage "reduction" procedure imposes that the grand-coalition satisfies superadditivity, in the sense that the grand coalition should be able to achieve in terms of aggregate worth at least the sum of what is achievable under any coalition structure (Ray and Vohra, 1996, 1997; Bloch, 1997). Indeed, whichever strategies chosen by the coalitions of any coalition structure can be replicated by the grand coalition. The superadditivity concerns just the grand coalition and it is not implied at the level of sub-coalitions, because of the presence of spillovers.

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

A.4. All players are ex-ante identical, which means that each player has the same strategy space in the second stage emission 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[5]. 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 $\pi = \{c_1, c_2, \ldots, c_m\}$, where $\sum_i c_i = n$. As a consequence, the payoff received by the players only depends on coalition sizes and not on the identities of the coalition members. The per-member partition function (partition function hereon) can thus be denoted by $p(k; \pi)$, which represents the payoff of a player belonging to the size-k coalition in the coalition structure $\pi$. Finally, let us denote by $\pi = \{a_{(r)}, \ldots\}$ r size-a coalitions in the coalition structure $\pi$.

---

[5] We consider the equal sharing rule as an assumption since it is not endogenously determined in the model. However Ray and Vohra (1996) provides a vindication for this assumption.

Recent developments of the theory of endogenous coalition formation (Cf. Bloch, 1997) have stressed the implications of allowing players to join different coalitions. However, in the case of global warming and climate change, the negotiating agenda focuses on a single agreement that countries have to decide whether or not to sign. Therefore, we assume that:

A.5. Countries are proposed to sign a single agreement. Hence, those which do not sign cannot propose a different agreement. From a game-theoretic viewpoint, this implies that only one coalition can be formed, the remaining defecting players playing as singletons.

Another important assumption concerns the behaviour of countries when they assess the impact on the other countries' action of their decision to sign (or not to sign) the agreement on climate change control. In the next section we will use the following assumption:

A.6. When defecting from a coalition $c_i$, each country assumes that the other countries belonging to $c_i$ remain in the coalition.[6]

This assumption, which is consistent with the so-called Nash conjectures, may be relaxed to take into account that countries strategic behaviour may be more complex, and that they may anticipate the reaction of the other countries to their own defection (or accession). Assumption A.6 will be relaxed in Section 4.

Finally, we need to introduce a technical assumption which is necessary to reduce the number of equilibria of the coalition formation game:

A.7. Above a minimum coalition size[7], each country's payoff function increases monotonically with respect to the coalition size (the number of signatories in the symmetric case).[8]

---

[6] This assumption is equivalent to the assumption of "Nash conjectures" in a simultaneous oligopoly game where a player assumes no change in the other players decision variable when it modifies its own decision variable. However, coalition theory often uses a different assumption, named "coalition unanimity" (Cf. Bloch, 1997), where the whole coalition is assumed to collapse when one of its members defects (see Chander and Tulkens, 1993, 1995). See the discussion in Section 4.

[7] The minimum size is equal to 2 in the case of orthogonal free-riding (see below). In this case, the individual payoff function increases for all feasible values of the coalition size. This may not be the case with non-orthogonal free-riding.

[8] This assumption excludes the possibility of "exclusive membership equilibria" where the group of cooperating countries can refuse entry to a country which wants to join the coalition (Yi, 1997).

This last assumption is quite natural in the case of negotiations on climate change control. Indeed, being climate a public good, each country which finds it convenient to reduce its own GHG emissions provides a positive contribution to the welfare of all countries (both inside and outside the coalition).

### 3. The coalition game

The first stage game consists of 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 $\pi = \{c, 1_{(n-c)}\}$. The distinctive features of the game describing negotiations on climate change control are as follows**:**

**-** Positive spillovers. In any single coalition structure, if some countries form a coalition, the other countries are better off. Then the partition function of any country 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 action. In particular, there may emerge two different free-riding behaviour patterns according to the slope of players' reaction functions (Carraro and Siniscalco, 1993).

**-** Orthogonal free-riding. When countries' reaction functions in the emission game are orthogonal, free riders just benefit from the cooperative action of the coalition and they cannot damage it (i.e. there is no "carbon leakage").

**-** Non orthogonal free-riding. When countries' reaction functions in the emission game are non-orthogonal, i.e. in an environment in which there is interdependence between countries' emission strategies, free-riders can damage the coalition by increasing their emissions whenever cooperating countries reduce their own. This implies that a small number of cooperators may loose from cooperation because of the increased emissions in the free-riding countries. Let us denote by $c^m$ the size a coalition has to reach in order to start benefiting from cooperation.[9]

In the case of climate change, both situations are feasible. Current estimates of "carbon leakage" range from very small values (Cf. OECD,1993) to quite large ones (Ulph, 1993). Therefore, it is worth analysing both

---

[9] Notice that $c^m$ does not necessarily coincide with the minimum size of assumption A.7 (called c^). An example is provided in Figure 3.

cases and the impact of different degrees of "carbon leakage" on the outcome of negotiations on climate change control.

Given these features of the n-country game, let us characterise its equilibrium. As said, countries decide independently, simultaneously, and non-cooperatively whether or not to sign the agreement on climate change control. Hence, the concept of Nash equilibrium (NE) is the most appropriate also for the first stage of the game – the coalition game. The Nash equilibrium is completely characterised by the following two properties, first derived in the cartel literature (D'Aspremont et al., 1983) and then often used also in the environmental literature (Carraro and Siniscalco, 1993, Barrett, 1994).

- <u>Profitability</u>. A coalition is profitable if each cooperating player gets a payoff larger than the one he would obtain in the autarchic state, i.e. when no coalition forms. Formally:

(1) $$p(c; \pi) \geq p(1; \pi^S),$$

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

From eq. (1), the value of the minimal profitable coalition size $c^m$ can be derived. 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 two (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 free-riders' action. This size is generally larger than two (see Figure 3).

- <u>Stability</u>. 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[10]. 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 players in the coalition c.[11] It is externally stable if no singleton is better off by joining the coalition c. Formally:

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

where $\pi=\{c,1_{(n-c)}\}$ and $\pi'= \pi\backslash\{c,1\}\cup\{c'\}$ and $c'= c+1$ for all players which do not belong to c.

Notice that assumption A.6 is implicit in this definition of stability, because each country chooses its strategy taking the action of the other countries as given. Then, each country, in order to define its (individually)

---

[10] Yi (1997) denotes this condition by the term "stand alone stable".

[11] We suppose that if a player is indifferent between joining the coalition or defecting, then he joins the coalition.

optimal strategy, compares for each conceivable c its payoff when it belongs to the coalition with its payoff when it decides not to sign the environmental agreement.

A useful tool to analyse the stability of the coalition, i.e. the Nash equilibrium of the coalition game, is the stability function L(c) proposed in Carraro and Siniscalco (1992). This function describes a country's incentive to join a coalition c:

(3) $L(c) = p(c;\pi) - p(1;\pi')$,

where $\pi=\{c,1_{n-c}\}$ and $\pi'= \{c',1_{n-c'}\}$ where $c'=c-1$

In order to understand the behaviour of the stability function, we need to describe the non-member partition function $p(1;\pi')$ and the member one $p(c;\pi)$. By the positive spillover property, the free-rider payoff function increases with the size of c, as shown in Figure 1. The member partition function $p(c;\pi)$ is drawn in Figure 2 for the case in which free-riding is orthogonal, whereas the same function with non-orthogonal free-riding is represented in Figure 3.

Moreover, positive spillovers imply that the stability function becomes negative at least when c = n[12]. Indeed, when n-1 countries reduce their emissions, the n-th one can enjoy a clean environment without paying any cost. The same reasoning can be applied to the n-1th country and so on. However, as shown by Proposition 1 below, this process does not necessarily lead to $\pi^s$, the singleton coalition structure. There may be indeed a non-trivial stable coalition structure. The size of the stable coalition is determined by conditions (2), i.e. by the largest integer c* smaller than c', where c' is defined by L(c') = 0. This size is the minimal externally stable and the maximal internally stable. The stability function is a decreasing function of c as in Figure 4. Note that c* is not necessarily greater than one.

Given the above assumptions and equilibrium concept, we can prove the following:

**Proposition 1**. The NE of a simultaneous single coalition game is the following coalition structure:

- $\pi^* = \{c^*,1_{(n-c^*)}\}$, when $c^m \leq c^*$;
- $\pi^S = \{1_n\}$, i.e. the singleton structure, when $c^m > c^*$.[13]

---

[12] Hoel and Schneider (1997) consider a convex stability function such that the formation of a coalition creates negative spillovers on non members (above a given coalition size). This implies that the stability function can be positive for c=n. Then, the problem of coalition formation turns out to be trivial, since the grand coalition is stable.

[13] The symmetry assumption A.4 implies that there is not a single Nash equilibrium of the game. The number of Nash equilibria depends on the number of players because, by symmetry, all sub-sets of N formed by c* players can be a NE of the game. We say that there is one Nash equilibrium because we refer to the typology of the equilibrium, rather than to the identity of countries in the coalition.

*Proof:* When $c^m \leq c^*$ the profitability condition is satisfied. Then by definition of c*, a single coalition structure in which c<c* is not a NE, because a singleton's best replay is to join the coalition. On the other hand, in a single coalition structure in which c>c* the best replay of a cooperating player is to defect. Thereby only when c = c* no country wants to change its strategy. The equilibrium coalition structure is thus $\{c^*, 1_{(n-c^*)}\}$.
When $c^m > c^*$, a coalition structure in which $c \geq c^m > c^*$ is not stable, because a cooperating player's best replay is to defect. Besides, in all single coalition structure in which $c < c^m$, by definition of $c^m$ all cooperating players have an incentive to leave the coalition since they gain a larger payoff in the non cooperative state. Then the only NE structure of the game is the non-cooperative outcome, i.e. the coalition structure $\pi^S$ with n singletons.
□

To understand the relevance of this result, consider Table 1, which 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 value in Table 1 are computed using the simple model proposed in Carraro and Siniscalco (1992) and then used by Barrett (1994), Chander and Tulkens (1995).

**Table 1. 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 |

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.
- *Fact 2*: if c < 3, a country's payoff is not lower when it belongs to the coalition than when it defects and the coalition becomes smaller.
- *Fact 3*: if c > 3, a country's payoff is larger when it does not belong to the coalition than when it joins it.

As a consequence, the NE of this game is π* = {3,1,1,1}. Consider indeed any coalition structure in which the coalition is larger than three. By Fact 3, the best replay 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 2 the best replay of one of the players outside the coalition is joining the coalition. Fact 1 guarantees that the stable coalition is also profitable.

Consider now the case in which there is carbon leakage, i.e. free-riding is non-orthogonal. The values of countries' payoff are summarised in Table 2.

**Table 2. 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 |

Again the incentive system is summarised by the following two facts:

- *Fact 1*: the profitability condition is satisfied for all c≥5
- *Fact 2*: for all c ≥2, a country's payoff is larger when it does not belong to the coalition than when it joins it.

As a consequence, the NE of this game is π* = {1,1,1,1,1,1}. Consider indeed any coalition structure in which the coalition is larger than one. By Fact 2, the best replay of any coalition member is to defect until the singleton structure is reached.

What are the implications of the above results? First, if an equilibrium coalition exists, as in the case of orthogonal free-riding, the outcome of the game is not the one in which no cooperation takes place (no countries sign the agreement) as it could be expected given the public good nature of the global environment. At the equilibrium there are instead two groups of countries, signatories and defectors, where the size of the group of signatories crucially depends on the slope of countries' reaction functions (the size of carbon leakage). Notice that the conclusion on the existence of an equilibrium coalition in which only a sub-group of countries cooperate is consistent with the present negotiating agenda whose goal is to achieve, at least as a first step, an agreement on climate change control only among Annex 1 countries.

However, the stable (and Pareto optimal because the payoff function increases monotonically) coalition is generally formed by a low number of players.[14] In the example of Table 1, at most 3 players, whatever n, join the stable coalition and only in the favourable case of orthogonal free-riding.

As a consequence of this last result, the recent literature on international environmental agreements (see Carraro, 1997a,b for a survey) focused on ways of broadening the "endogenous" stable coalition by "exogenously" introducing appropriate additional policy measures. Three ideas deserve our attention.

- <u>Transfers</u>

Transfers are often proposed to tackle the profitability dimension of international negotiations, i.e. to compensate those countries which, because of their asymmetries, would lose from signing the agreement. Transfers may also be an important tool to expand an originally stable, but small, environmental coalition. However, as shown in Carraro and Siniscalco (1993), countries which accept to implement a transfer program to non-signatories must be <u>committed</u> to cooperation (this condition is weaker with asymmetric countries; see Botteon and Carraro, 1997a). As a consequence the international agreement becomes partially self-enforcing.

- <u>Issue linkage</u>

As for transfers, the linkage of environmental negotiations to other economic issues (e.g. trade, technological cooperation) may be useful: (a) to reduce the constraints that asymmetries impose on the emergence of stable environmental agreements;[15] (b) to increase the size of the stable coalition. This second objective can be achieved even when all countries gain from signing the agreement if issue linkage is designed to offset

[14] This conclusion depends on the assumption on the returns from abatement activities and on the presence of abatement fixed costs (Barrett, 1994; Heal, 1994).

[15] This point was made by Folmer et al (1993) and Cesar and De Zeeuw (1996).

countries' free-riding incentives (Cf. Carraro and Siniscalco, 1995). This is the case when the negotiation on an issue with excludable benefits (a club good) is linked to the environmental negotiation (which, if successful, typically provides a public good, i.e. a non-excludable benefit). An example could be the linkage of environmental negotiations with negotiations on technological cooperation whose benefits are largely shared among the signatories whenever innovation spillovers to non-signatories are low (Cf. Carraro and Siniscalco, 1997).[16]

- Threats

The number of signatories of an international environmental agreement could be increased were non signatories threatened to be punished through adequate economic (e.g. trade) sanctions (Cf. Barrett, 1995, 1997b). However, credible threats are difficult to design. Emissions themselves are hardly a credible threat, because countries are unlikely to sustain self-damaging policies (e.g. when the "social clauses" of GATT are violated). Moreover, in this case, asymmetries play a double role. On the one hand, some countries may not gain from signing the environmental agreement; on the other hand, some countries, even when gaining from environmental cooperation, may lose from carrying out the economic sanctions. This may reduce the effectiveness of threats in increasing the number of signatories of international environmental agreements.

## 4. A rational coalitional behaviour

There is an extension which so far has not been considered in the literature. This extension concerns the behaviour of countries when they assess the impact of their decision of leaving (or joining) the environmental coalition.

In the environmental game analysed in Section 3 we adopted the usual Nash assumption, i.e. in both stages of the game, each country takes the other countries' decision as given when determining its optimal strategy. This implies that even in the first stage, when a country decides whether or not to join a coalition, it takes the other country's strategy as given. This assumption has been criticised by several authors (Cf. Bloch, 1997; Chander-Tulkens, 1995; Chew, 1994) who have also proposed some alternatives. Take for example the proposal -- already mentioned in footnote 6 -- of assuming coalition unanimity. Here the idea is that no coalition can form without the unanimous consensus of its members. In other words, a country knows that its decision to free-ride will lead the whole coalition to collapse.

[16] An extension to the case of structurally asymmetric countries is provided in Botteon and Carraro (1997b), whereas information asymmetries are accounted for in Katsoulacos (1997).

As a consequence of this coalitional behaviour assumption, the NE of the game is computed by comparing the payoff of each country when it belongs to the coalition with its payoff when it defects, for each conceivable subset of players in the coalition. However, when a country defects, the coalition breaks up into singletons. Thus, the payoff of a defecting country is the non cooperative one. Formally, a coalition can emerge if $p(c; \pi) > p(1; \pi^S)$, where $\pi=\{c,1_{(n-c)}\}$ and $\pi^S = \{1_n\}$, for all $i \in c$. Notice that this is the profitability condition previously defined.

As a consequence the equilibrium conditions are as follows. A coalition is stable iff it is:

- internally stable, i.e. profitable, because of coalition unanimity;
- externally stable, i.e. $p(1; \pi) > p(c'; \pi')$ for all $i \notin c$, where $\pi=\{c,1_{(n-c)}\}$ and $\pi' = \pi \backslash \{c,1\} \cup \{c'\}$ and $c' = c+1$.

These conditions are clearly less stringent than conditions (2) previously provided. As a consequence, the equilibrium stable coalition is generally larger, as shown by the following proposition.

**Proposition 2**. The NE of a simultaneous single coalition game under the coalition unanimity rule is the following coalition structure:

- when $c^m > c^*$, $\{1_{(n)}\}$ for all $c<c^m$, and $\{c,1_{(n-c)}\}$ for all $c \geq c^m$;
- when $c^m \leq c^*$, $\{c^*,1_{(n-c^*)}\}$ for all $c \leq c^*$, $\{c,1_{(n-c)}\}$ for all $c>c^*$

Furthermore, all coalition structures in which $c \geq c^*$ and $c \geq c^m$ are Pareto-optimal.

*Proof:* Consider the first part of the proposition. If $c<c^m$, it is straightforward that the singleton structure will emerge. On the other hand, each coalition structure in which $c \geq c^m$ is stable, as no singleton wants to join the coalition, because $c^*<c^m$, and no cooperating country wants to leave the coalition, because its defection would induce the singleton structure and by definition of $c^m$ it would be worse off.

Consider now the second part of the proposition. In each coalition structure in which $c<c^*$ a singleton wants to join the coalition until the size $c^*$ is reached. By monotonicity of the payoff function, all players are better off and the structure $\pi^* = \{c^*,1_{(n-c^*)}\}$ will be implemented. Moreover, all coalitions in which $c \geq c^*$ are also stable, because no singleton wants to join the coalition and no cooperating player has an incentive to leave the coalition because he would induce the singleton structure.

Finally, all coalitions satisfying $c \geq c^*$ and $c \geq c^m$ are Pareto optimal. Given a coalition of size c, it is indeed impossible to modify its size without reducing the payoff of at least one country. If the size is reduced, monotonicity (assumption A.7) implies that the payoff of coalition members is reduced. If the size is increased, coalition members increase their payoff, but the payoff of the country joining the coalition is reduced (by definition of $c^*$). □

The problem with the idea of coalition unanimity is that the implied countries' strategy is not credible. If, as shown by Proposition 1, countries have an incentive to form a coalition, why should the idea of moving to the non-cooperative equilibrium -- as a reaction to a defection from the coalition -- be credible?

Moreover, both the Nash assumption of Section 3 and coalition unanimity turn out to be exogenous assumptions. Hence, a better approach could be the one in which the actual coalition's reaction to a deviation is accounted for, without referring to a fixed behavioural pattern as in the previous games. This issue can be tackled by using a different solution concept. Let us introduce the idea of rational conjectures, i.e. the idea that a country accounts for the other countries' anticipated reactions when it establishes its strategy. In a game of coalition formation, this means that each player, before defecting, predicts what would happen after his deviation, i.e. how many other cooperating players will leave the coalition, and how many singletons will join the coalition.

In this context, we need to re-define the equilibrium 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. A coalition is:

- internally stable, i.e. no cooperating player would be better off in the coalition structure induced by his defection iff $p(c;\pi) \geq p(1;\pi')$ 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);
- externally stable, i.e. no singleton would be better off in the coalition structure induced by his accession to the coalition iff $p(1;\pi) \geq p(c''; \pi'')$ where $\pi'' = \pi \backslash \{c, 1_{(k'+1)}\} \cup \{c''\}$ and $c''=c+k'+1$, where k' is the number of accessions following i's accession (k'+1 accessions including i).

These definitions imply that, in the coalition game with rational conjectures, a country compares its payoff in the initial coalition structure with its payoff in the final equilibrium structure, where the latter is the coalition structure which actually emerges after the country's defection or accession.

The equilibrium outcome depends on the initial coalition structure, as shown by the following proposition.

**Proposition 3**. In a game with rational conjectures, the equilibrium coalition structures are defined as follows.

- when $c^m \leq c^*$, if the initial coalition is $c \leq c^*$, the stable coalition structure is $\pi^* = \{c^*, 1_{(n-c^*)}\}$, whereas if $c > c^*$, the stable coalition structure is one of the *contingent* structures $\pi^{**} = \{c^{**}, 1_{(n-c^{**})}\}$ in which $c^{**} \leq c$.
- when $c^m > c^*$, if the initial coalition structure is $c < c^m$, the stable coalition structure is the singleton one, whereas if $c \geq c^m$, the stable coalition structure is $\pi^m = \{c^m, 1_{(n-cm)}\}$.

*Proof*: Consider the first part of the proposition. If $c<c^*$, by external stability singletons have an incentive to join the coalition until the structure $\pi^*= \{c^*, 1_{(n-c^*)}\}$ is implemented and once such a structure is implemented no player has an incentive to move. Then, the rational conjecture for a player belonging to c is k=0 (no defections) and k'=(c*-c+1) (k' accessions in order to achieve a number of signatories equal to c*). The stable coalition structure is $\pi^*$. Similarly, if c=c*, a country belonging to c anticipates that no defection occurs (k=0), because c* is stable, and that if it defects it will be replaced (k'=1). The equilibrium coalition structure is again $\pi^*$.

If c>c* a cooperating country has an incentive to defect but before defecting it considers to what outcome its deviation would lead. For instance, if c=c*+1, once a country belonging to c defects, the minimal stable size is reached and no other player has an incentive to move. Thereby, the rational conjecture is k=0 and k'=0. The stable coalition structure is again $\pi^*$.

It is clear that the dependence of the equilibrium coalition structure on the initial coalition makes it difficult to provide a general result. For instance, when c=c*+2, a country's defection leads to the coalition structure in which c=c*+1. This coalition structure is not stable, as previously discussed, and one more player will defect until $\pi^*$ is reached. Hence, the rational conjecture is k=1 and k'=0. Therefore, before defecting, the cooperating country compares $p(c; \pi)$ -- its payoff in the initial coalition structure -- with $p(1; \pi^*)$ -- the payoff to which its defection would lead. If $p(c; \pi) \geq p(1; \pi^*)$, it will not defect, and the coalition c=c*+2 would be stable.

This type of analysis has to be performed for all coalition structures in which c>c*. This implies that more than one coalition structure may be stable in this game (see the example below). Let us denote by $\pi^{**}= \{c^{**}, 1_{(n-c^{**})}\}$ the stable coalition structures of this game and let us call them <u>contingent structures</u>, because they depend on the initial coalition size and on the specific payoff function. According to the initial coalition size c, a particular contingent structure is stable, but for each c there is one and no more than one corresponding contingent coalition structure. Then, c** = c* is just one of the coalition structures belonging to the stable contingent coalition structure set of this game. In particular, it is the smallest one.

Consider now the second part of the proposition. If the initial structure is such that $c<c^m$, the profitability condition is not satisfied. Then, the rational conjecture is k=c-1 and k'=0. The stable coalition structure is the singleton one. If $c=c^m$, a player's defection is not followed by a new accession (k'=0), but rather by further defections until the singleton structure is reached ($k=c^m-1$). Hence, no player will defect from $c=c^m$. When $c>c^m$ , if a player defects other deviations follow until the coalition size $c^m$ is reached. Once it is reached, no player defects, because after his defection the coalition would break up into singletons. Hence, the rational conjecture is $k=c-c^m-1$ and k'=0. The stable coalition structure is $\{c^m, 1_{(n-cm)}\}$. □

Note that only in the case $c^*<c^m$, the stability of the final coalition structure hinges on the threat of the implementation of the singleton structure (like under the assumption of coalition unanimity), whereas in the case $c^*>c^m$, this threat is not credible. There is also an interesting corollary of Proposition 3.

**Corollary 1**. The grand coalition is stable if the payoff of a player in the grand coalition is larger than the payoff he receives as a singleton free-rider in the coalition structure with the largest $c^{**}<n$.

The examples of climate change negotiations summarised by Tables 1 and 2 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 1). The starting point are again the set of Facts 1 to 3.

If the initial 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 implemented. 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 initial 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 implemented. Since $\pi^*$ 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. Then, $\pi'$ is not a stable structure (the stable structure is again $\pi^*$).

If the initial structure is $\pi''=\{1,5\}$, again a cooperating country has an incentive to defect. If it defects, the structure $\pi'$ is implemented; 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 accessions to replace the two defectors). The consequence is that no country belonging to c=5 decides to defect because $p(5;\pi'') > p(1;\pi^*)$. Thereby, the structure $\pi''$ is stable.

If the initial structure is $\pi'''=\{6\}$, a cooperating country is willing to defect. If it defects, the stable structure $\pi''=\{1,5\}$ is implemented. 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 adopting rational conjectures, the stable outcome of the game summarised by Table 1 is either the basic coalition structure $\pi^*=\{1,1,1,3\}$ or the contingent coalition structure $\pi''=\{1,5\}$.

Consider now Table 2, where the game with non-orthogonal free-riding is summarised. If the initial 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). Therefore, the stable outcome is the singleton structure.

If the initial 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 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 its partition function $p(5; \pi'')$ with its partition function in the singleton structure $p(1; \pi^s)$. As a consequence, the coalition structure $\pi''$ is stable.

Finally, if the initial 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 $\pi''$ (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 non-orthogonal free-riding with rational conjectures there are again two stable coalition structures: the trivial coalition structure $\pi^s=\{1,1,1,1,1,1\}$ and the contingent structure $\pi'' = \{1,5\}$.

## 5. Conclusions

The analysis of the game which captures the incentive scheme which is behind current negotiations on climate change control enabled us to achieve some relevant conclusions.

- In contrast with results obtained in 2-player games, cooperation is a possible equilibrium outcome of the n-player game. However, only a subset of players is likely to sign the cooperative agreement.

- This subset is larger the lower “carbon leakage”, i.e. the interdependency of countries’ emission strategy, but is likely to be insufficient to mitigate the effects of carbon emissions on climate.

- A larger equilibrium coalition emerges when countries correctly conjecture the consequences of their decision to defect from the coalition (or to join it).

There are several directions of further research that deserve additional efforts. The strategic dimension of environmental negotiations, both at the international and domestic levels (voters may be asked to ratify and environmental agreement) opens interesting political economy problems (Currarini and Tulkens, 1997; Carraro and Siniscalco, 1998). The lack of a supra-national authority calls for an analysis of new international institutions (Compte and Jehiel, 1997 propose an international arbitrator). The possibility to expand coalitions by linking environmental and trade negotiations requires further theoretical and empirical analyses. A dynamic framework may be more appropriate to deal with environmental issues in which the stock of pollutants, rather than the flow (emissions) is the crucial variable to monitor (Cf. Van der Ploeg and De Zeeuw, 1992). Theoretical and simulation analyses with heterogeneous countries would also be very important to understand which factors influence a country's decision to sign (or not sign) an international agreement on climate change (first attempts are contained in Barrett, 1997a; Botteon and Carraro, 1997b).

## References


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

Barrett, S. (1995), “Trade Restrictions in International Environmental Agreements”, mimeo, London Business School.

Barrett, S. (1997a), "Heterogeneous International Environmental Agreements", in C. Carraro, ed., International Environmental Agreements: Strategic Policy Issues, E. Elgar, Cheltenham.

Barrett, S. (1997b), "Towards a Theory of International Cooperation" in C. Carraro and D. Siniscalco, eds., New Directions in the Economic Theory of the Environment, Cambridge University Press, Cambridge.

Bloch, F. (1997), “Non-cooperative 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 Agreements: Strategic Policy Issues, E. Elgar, Cheltenham.

Botteon, M. and C. Carraro (1997b), "Strategies for Environmental Negotiations: Issue Linkage with Heterogeneous Countries", in H. Folmer and N. Hanley, eds., Game Theory and the Global Environment, E. Elgar, Cheltenham.

Carraro, C. (1997a), ed., International Environmental Agreements: Strategic Policy Issues, E. Elgar, Cheltenham.

Carraro, C. (1997b), “Environmental Conflict, Bargaining and Cooperation” forthcoming in J. van den Bergh, Handbook of Natural Resources and the Environment, E. Elgar, Cheltenham.

Carraro, C. and D. Siniscalco (1992), "The International Protection of the Environment: Voluntary Agreements among Sovereign Countries", forthcoming in Dasgupta, P and K.G. Maler, The Economics of Transnational Commons, Clarendon 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 (1995), "Policy Coordination for Sustainability: Commitments, Transfers, and Linked Negotiations", in I. Goldin and A. Winters, eds., The Economics of Sustainable Development, Cambridge University Press, Cambridge.

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.

Carraro, C. and D. Siniscalco (1998), “International Environmental Agreements. Incentives and Political Economy”, forthcoming in the *European Economic Review*.

Cesar, H. and A. De Zeeuw (1996), "Issue Linkage in Global Environmental Problems", in A. Xepapadeas, ed., Economic Policy for the Environment and Natural Resources, E. Elgar, Cheltenham.

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

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, Milan.

Chew, M.S. (1994), "Farsighted Coalitional Stability", *Journal of Economic Theory*, 54, 299-325.

Compte, O. and P. Jehiel (1997), "International Negotiations and Dispute Resolution Mechanisms: the Case of Environmental Negotiations", in C. Carraro, ed., International Environmental Agreements: Strategic Policy Issues, E. Elgar, Cheltenham.

Currarini, S. and H. Tulkens (1997), "Core-Theoretic and Political Stability of International Agreements on Transfrontier Pollution", CORE Discussion Paper, Louvain-la-Neuve.

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.

Folmer, H., van Mouche, P. and S. Ragland (1993), "Interconnected Games and International Environmental Problems", *Environmental Resource Economics*, 3, 313-335.

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

Hoel, M. and K. Schneider (1997), "Incentives to Participate in an International Environmental Agreement", *Environmental Resource Economics*, 9, 153-170.

Katsoulacos, Y. (1997), R&D Spillovers, R&D Cooperation, Innovation and International Environmental Agreements, in C. Carraro, ed., International Environmental Agreements: Strategic Policy Issues, E. Elgar, Cheltenham.

Mariotti, M. (1997), "A Model of Agreements in Strategic Form Games", *Journal of Economic Theory*, 73, 128-139.

OECD (1993), The Cost of Cutting Carbon Emissions: Results from Global Models, Paris.

Ray, D. and R. Vohra (1996), "A Theory of Endogenous Coalition Structure", mimeo, Dept. of Economics, Brown University and Boston University.

Ray, D. and R. Vohra (1997), "Equilibrium Binding Agreements", *Journal of Economic Theory*, 73, 30-78.

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

Van der Ploeg, F. and A.J. De Zeeuw (1992), "International Aspects of Pollution Control", *Environmental and Resource Economics*, 3, 117-139.

Yi, S. (1997), "Stable Coalition Structures with Externalities", forthcoming, *Games and Economic Behaviour*.

FIGURE 1. **Non member partition function**

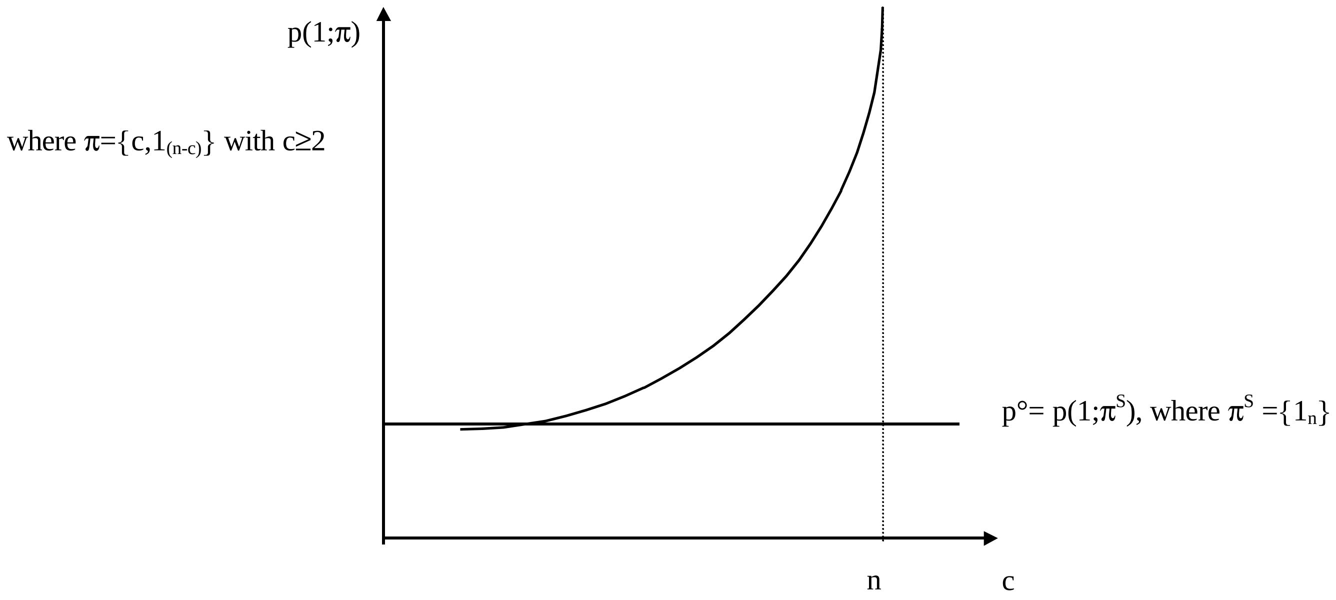


FIGURE 2. **Member partition function with orthogonal free-riding**

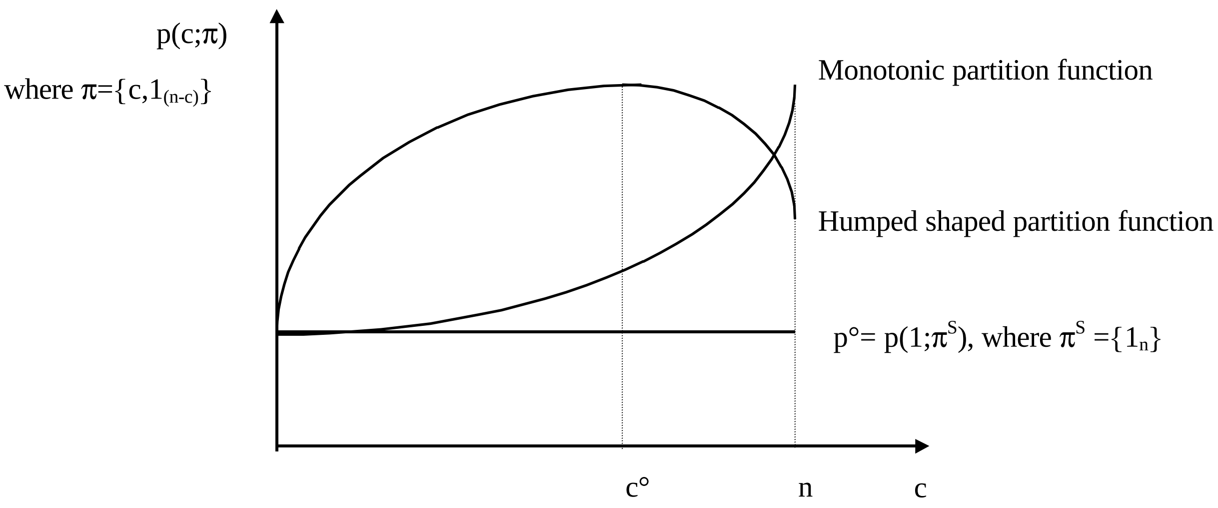

FIGURE 3. **Member partition function with non-orthogonal free-riding**

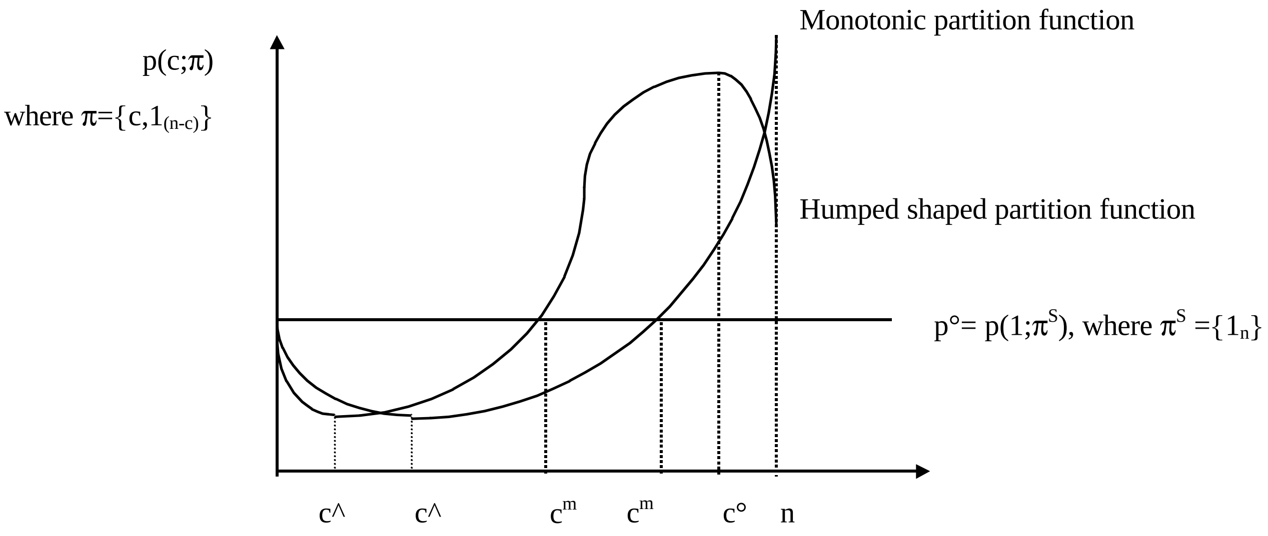


FIGURE 4. **Stability function**